Probabilistic Exploitation of the Lucas and Kanade Smoothness Constraint

Volker Willert, Julian Eggert, Mark Toussaint, Edgar Körner

2008

Preprint:

This is an accepted article published in Machine Learning and Applications. The final authenticated version is available online at: https://doi.org/[DOI not available]

Probabilistic Exploitation of the Lucas and Kanade Smoothness Constraint

Volker Willert, Julian Eggert, Marc Toussaint*, Edgar Körner
HRI Europe GmbH — TU Berlin*
Carl-Legien-Str. 30 — Franklinstr. 28/29
D-63073 Offenbach, Germany — D-10587 Berlin, Germany
volker.willert@honda-ri.de — mtoussai@cs.tu-berlin.de

Abstract

The basic idea of Lucas and Kanade is to constrain the local motion measurement by assuming a constant velocity within a spatial neighborhood. We reformulate this spatial constraint in a probabilistic way assuming Gaussian distributed uncertainty in spatial identification of velocity measurements and extend this idea to scale and time dimensions. Thus, we are able to combine uncertain velocity measurements observed at different image scales and positions over time. We arrive at a new recurrent optical flow filter formulated in a Dynamic Bayesian Network applying suitable factorisation assumptions and approximate inference techniques. The introduction of spatial uncertainty allows for a dynamic and spatially adaptive tuning of the constraining neighborhood. Here, we realize this tuning dependent on the local Structure Tensor of the intensity patterns of the image sequence. We demonstrate that a probabilistic combination of spatiotemporal integration and modulation of a purely local integration area improves the Lucas and Kanade estimation.

1. Introduction

The local and linear differential method of Lucas and Kanade [9] is one of the most popular approaches for optical flow computation. Compared to global smoothness constraints used for example by Horn and Schunk [8], their local explicit method is more accurate and more robust with respect to errors in gradient measurements [2]. Nevertheless, the local approach suffers from the aperture problem and the linearisation of the underlying constancy assumption for image intensity. Lucas' and Kanade's basic idea to assume that the optical flow field is spatially constant within some neighborhood is in many cases not enough to resolve motion ambiguities and does in particular not hold at motion boundaries. Further on, the linearised intensity constancy assumption is suitable only for small displacements.

In this paper, we reduce the above mentioned drawbacks just by exploiting the local Lucas-Kanade constraint consistently but without introducing further global smoothness constraints. The main assumption of Lucas and Kanade is that a number of pixels within a neighborhood move with the same velocity. This implies that the movement of every single pixel is influenced by the movement of its neighboring pixels. So one could argue that the movement of an observed pixel is equated with the movement of an image patch centered around this pixel. Since this is not true, one could follow a slightly different implication and assume that the position of a pixel is uncertain and so its velocity can only be inferred by taking the neighborhood into account. Either way, the question arises what is an appropriate size of the neighborhood and which neighboring position influences the pixel velocity how strongly?

Usually, the neighborhood is weighted with a Gaussian assuming a less likely contribution to the velocity estimate of the center pixel for larger distances to the center [2]. More elaborate approaches adapt the neighboring influence by taking the underlying structure into account [4], [6] which improves the optical flow accuracy, especially at motion boundaries.

Keeping in mind the above mentioned considerations we start with a probabilistic interpretation of the Lucas-Kanade approach inspired by the work of Simoncelli *et al.* [12]. We propose a *generative model* that allows to infer the velocity of every pixel from the movement of image patches. Every pixel within the patch is assigned a different uncertainty to be able to adaptively adjust the neighborhood influence on the velocity estimate. This leads to an observation likelihood for optical flow estimation described in section 2.1.

Another important aspect of motion estimation is the fact that motion is a dynamic feature of an image sequence. Thus, the longer the spatiotemporal process is observed the more precisely we can estimate and predict the motion contained in an image sequence. This has motivated several approaches [13], [3], [7] to recursively estimate the optical flow over time including a prediction model that defines

some temporal relation between pixel movements. Along this line of argumentation, it seems straightforward to also include a local constraint on the pixel movements within some neighborhood in time.

Here, we realize this idea and extend the Lucas-Kanade constraint to the time dimension. More precisely, we assume that the pixel velocities within a spatiotemporal neighborhood remain constant. This constraint is included via a temporal transition that consists of factors which are dependent on neighboring positions. The influence of these neighboring positions is again allowed to be spatiotemporally adaptive. In the same way, we account for the limitations of the linear differential method and incorporate a probabilistic coarse-to-fine strategy inspired by [11]. Our method propagates motion information over scales via a scale-transition consisting of factors which are adaptively dependent on the movement of pixels from neighboring scales (for details see section 2.2). Combining both transitions we obtain a new recurrent *scale-time filter* (STF) for optical flow estimation that incorporates the idea of the local Lucas-Kanade constraint to both dimensions, scale and time. This is formulated in a Dynamic Bayesian Network (DBN) and described in section 2.3.

To enhance optical flow performance at motion discontinuities, we allow for adaptation of the neighborhood uncertainty dependent on structural information. This is related to discontinuity preserving anisotropic diffusion approaches [4]. Here, the adaptation of the neighborhood influence is realized with adaptive Gaussian kernels whereas the orientation and the sharpness of the Gaussians are gained from the structural information provided by the local Structure Tensor of the underlying intensity pattern corresponding to the neighborhood (see section 3).

In section 4 we compare different types of realizations of the proposed STF with state-of-the-art algorithms that achieve highest accuracy in optical flow computation, like the CLG approach [5], to shortly discuss the advantages of recurrent filtering techniques in contrast to variational approaches that do not propagate motion information via predictive models.

2 Probabilistic Recurrent Filter

To describe the filtering process a DBN as depicted in Fig. 1 is proposed. We assume a generative model for the observables $\mathbf{Y}^{tk}$ of an image sequence $\mathbf{I}^{1:T,1:K}$ with T images at equidistant points in time t at K spatial resolution scales k with $t' = t + 1$ and $k' = k + 1$ being the next time step and the next finer scale, respectively.

Here, the observable $\mathbf{Y}^{tk} = (\nabla\mathbf{I}^{tk}, \mathbf{I}_t^{tk}) \in \mathbb{R}^{3\times X^k}$ is defined over the image range X^k of the particular scale k and comprises the gradient field $\nabla\mathbf{I}^{tk} \in \mathbb{R}^{2\times X^k}$ with entries $\nabla I_{\mathbf{x}}^{tk} = (I_{x\mathbf{x}}^{tk}, I_{y\mathbf{x}}^{tk})^T \in \mathbb{R}^{2\times 1}$ and the scalar field of

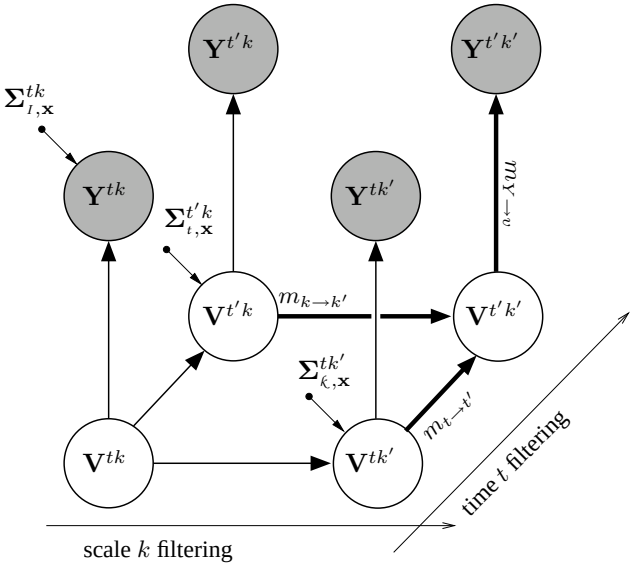

Figure 1. Dynamic Bayesian Network for recurrent motion estimation realizing a scale-time filter (STF) that simultaneously propagates beliefs along scales k and time t.

temporal derivatives $\mathbf{I}_t^{tk} \in \mathbb{R}^{X^k\times 1}$ with entries $I_{t,\mathbf{x}}^{tk} \in \mathbb{R}^{1\times 1}$ at each pixel location $\mathbf{x} \in \mathbb{N}^{X^k}$ of the image at a particular time t and scale k. In the following, the gradient field is assumed to be deterministic. Furthermore, we define a scalar field patch of temporal derivatives centered around $\mathbf{x}$ as $\mathbf{I}_{t,\mathbf{x}}^{tk} \in \mathbb{R}^{X^k\times 1}$ (which should not be confused with $I_{t,\mathbf{x}}^{tk}$) and a gradient field patch centered around $\mathbf{x}$ as $\nabla\mathbf{I}_{\mathbf{x}}^{tk} \in \mathbb{R}^{2\times X^k}$. Similarly, the hidden state $\mathbf{V}^{tk} \in \mathbb{R}^{2\times X^k}$ is a flow field at time slice t and scale k defined over the image range X^k with entries $\mathbf{v}_{\mathbf{x}}^{tk} \in \mathbb{R}^{2\times 1}$ at each pixel location $\mathbf{x}$ of the image.

The probabilistic generative model is precisely defined by the specification of the *observation likelihood* for the image derivates $\mathbf{Y}^{tk}$ formulated in (1) and the *transition probability* for the flow field $\mathbf{V}^{t'k'}$ at the new timestep t' at finer scale k' specified in (2) given the flow field $\mathbf{V}^{t'k}$ at the same time t' but coarser scale k and the flow field $\mathbf{V}^{tk'}$ from last time t but at the same scale k'.

For the observation likelihood and the flow field transition probability we assume that they factorise over the image as follows

$$P(\mathbf{Y}^{tk} \mid \mathbf{V}^{tk}) = \prod_{\mathbf{x}} \ell(\mathbf{v}_{\mathbf{x}}^{tk}) \,, \tag{1}$$

$$P(\mathbf{V}^{t'k'} \mid \mathbf{V}^{t'k}, \mathbf{V}^{tk'}) = \prod_{\mathbf{x}} \phi_{\ell}(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) \times \phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) \,. \tag{2}$$

Equation (2) explicitly expresses that the conditional dependence $P(\mathbf{v}_{\mathbf{x}}^{t'k'} \mid \mathbf{V}^{t'k}, \mathbf{V}^{tk'})$ can be split in two pairwise potentials ϕ_k, ϕ_t, as explained in detail in section 2.2. This will allow us to maintain only factored beliefs during inference, which makes the approach computationally practicable.

2.1 Observation likelihood

We follow a similar argumentation as Simoncelli *et al.* [12] to obtain the $\ell(\mathbf{v}_{\mathbf{x}}^{tk})$-factors (1) of the observation likelihood. However, our likelihood results from a generative model assuming that a scalar field patch of temporal derivatives $\mathbf{I}_{t,\mathbf{x}}^{tk} \in \mathbb{R}^{X^k \times 1}$ centered around $\mathbf{x}$ is generated by the velocity $\mathbf{v}_{\mathbf{x}}^{tk} \in \mathbb{R}^{2\times 1}$ at position $\mathbf{x}$ and the gradient field patch $(\nabla \mathbf{I}_{\mathbf{x}}^{tk})^T \in \mathbb{R}^{X^k \times 2}$ centered around the same position $\mathbf{x}$.

While introducing this model based on *patches* around position $\mathbf{x}$ instead of only the *pixel* at position $\mathbf{x}$ itself we imply that the optical flow is locally constant in a sense similar to the Lucas-Kanade constraint [9]. Additionally, we assume i.i.d. additive Gaussian noise s_t, $\mathbf{S}_v$ on the temporal derivatives and the flow field, respectively.

$$\ell(\mathbf{v}_{\mathbf{x}}^{tk}) = \mathcal{N}(-\mathbf{I}_{t,\mathbf{x}}^{tk} \mid (\nabla \mathbf{I}_{\mathbf{x}}^{tk})^T \mathbf{v}_{\mathbf{x}}^{tk}, \mathbf{\Sigma}_{\ell,\mathbf{x}}^{tk}) \,, \tag{3}$$

$$\mathbf{\Sigma}_{\ell,\mathbf{x}}^{tk} = \begin{pmatrix} \ddots & \dots & \mathbf{0} \\ \vdots & \sigma_{\ell,\mathbf{x}\mathbf{x}'}^{tk} & \vdots \\ \mathbf{0} & \dots & \ddots \end{pmatrix} \,, \tag{4}$$

$$\sigma_{\ell,\mathbf{x}\mathbf{x}'}^{tk} = \frac{(\nabla I_{\mathbf{x}'}^{tk})^T \mathbf{S}_v \nabla I_{\mathbf{x}'}^{tk} + s_t}{f_\ell(\mathbf{x}', \mathbf{x}, t, k)} \,. \tag{5}$$

In notation (3), the patches can be regarded as vectors and the covariance matrix $\mathbf{\Sigma}_{\ell,\mathbf{x}}^{tk}$ is a diagonal with entries $\sigma_{\ell,\mathbf{x}\mathbf{x}'}^{tk}$ that depend on the position $\mathbf{x}'$ relative to the center $\mathbf{x}$, the time t, the scale k, the flow field covariance $\mathbf{S}_v$ and the variance on the temporal derivatives s_t. Here, f_ℓ takes into account the spatial uncertainty of the velocity measurement and can implement any kind of spatial weighting, such as a binomial blurring filter proposed in [11] or an anisotropic and inhomogenous Gaussian weighting $f_\ell = \mathcal{N}(\mathbf{x}'|\mathbf{x}, \mathbf{\Sigma}_{I,\mathbf{x}}^{tk})$ which we investigate in section 3.

In contrast to [11], we introduced time t as an additional dimension and derived a more compact notation by putting the spatial weighted averaging directly into the likelihood formulation defining multivariate Gaussian distributions for vectors that describe image patches centered around image locations. Allowing for uncertainties $\mathbf{\Sigma}_{\ell,\mathbf{x}}^{tk}$ that are adaptive in location $\mathbf{x}$, scale k and time t we are able to tune the local motion measurements dynamically dependent on the underlying structure of the intensity patterns as explained in section 3.

2.2 Flow field transition probability

The flow field transition probability defined by equation (2) consists of two pairwise potentials. The first potential $\phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'})$ assumes that the flow field at every spatial scale k transforms from $t \rightarrow t'$ according to itself. The second potential $\phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k})$ realizes a refinement from coarser to finer scale $k \rightarrow k'$ at every time t' via interpolation of the coarser flow $\mathbf{V}^{t'k}$.

To motivate the first transition factor we assume that the origin of a local flow vector $\mathbf{v}_{\mathbf{x}}^{t'k'}$ at position $\mathbf{x}$ at time t' was a previous flow vector $\mathbf{v}_{\mathbf{x}'}^{tk'}$ at some corresponding position $\mathbf{x}'$ at time t,

$$\mathbf{v}_{\mathbf{x}}^{t'k'} \sim \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}'}^{tk'}, \sigma_t) \,, \tag{6}$$

which says that the change in time of the flow field is white with undirectional transition noise between $\mathbf{V}^{tk'}$ and $\mathbf{V}^{t'k'}$. Now, asking what the corresponding position $\mathbf{x}'$ in the previous image was, we assume that we can infer it from the flow field itself as follows

$$\mathbf{x}' \sim f_t(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}) := \mathcal{N}(\mathbf{x}'|\mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{\Sigma}_{t,\mathbf{x}}^{tk}) \,. \tag{7}$$

In principle f_t can be any arbitrary function. Here, we define it as an inhomogeneous anisotropic Gaussian to be able to steer the orientation and to adapt the strength of the uncertainty in spatial identification $\mathbf{\Sigma}_{t,\mathbf{x}}^{tk}$ between corresponding positions in time (see section 3 for details). Note that here we use $\mathbf{v}_{\mathbf{x}}^{t'k'}$ to retrieve the previous corresponding point $\mathbf{x}'$ which is a suitable approximation keeping in mind that we have assumed directly beforehand $\mathbf{v}_{\mathbf{x}}^{t'k'} \approx \mathbf{v}_{\mathbf{x}'}^{tk'}$ in (6). Combining both factors (6) and (7) and integrating $\mathbf{x}'$ we get the first pairwise potential

$$\phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) = \sum_{\mathbf{x}'} \mathcal{N}(\mathbf{x}'|\mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{\Sigma}_{t,\mathbf{x}}^{tk}) \times \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}'}^{tk'}, \sigma_t) \,, \tag{8}$$

that imposes a spatial coherence constraint on the flow field combined with a linear stochastic drift. Equivalent to (6) for the second transition factor we assume that the origin of a local flow vector $\mathbf{v}_{\mathbf{x}}^{t'k'}$ at position $\mathbf{x}$ at finer scale k' corresponds to a flow vector $\mathbf{v}_{\mathbf{x}''}^{t'k}$ from coarser scale k at some corresponding position $\mathbf{x}''$,

$$\mathbf{v}_{\mathbf{x}}^{t'k'} \sim \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}''}^{t'k}, \sigma_k) \,, \tag{9}$$

assuming white transition noise σ_k. Since it is uncertain how strong a position $\mathbf{x}''$ at coarser scale k influences the velocity estimate at position $\mathbf{x}$ at finer scale k', we assume that we can infer it from the neighborhood similar to (7)

$$\mathbf{x}'' \sim f_k(\mathbf{x}'', \mathbf{x}) := \mathcal{N}(\mathbf{x}''|\mathbf{x}, \mathbf{\Sigma}_{k,\mathbf{x}}^{tk}) \,. \tag{10}$$

For the same reasons as mentioned for the temporal transition factor (8) we choose f_ℓ to be also an adaptive Gaussian kernel. Again, combining both factors (9) and (10) and integrating $\mathbf{x}''$ we get the second pairwise potential

$$\phi_\ell(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) = \sum_{\mathbf{x}''} \mathcal{N}(\mathbf{x}''|\mathbf{x}, \mathbf{\Sigma}_{\ell,\mathbf{x}}^{tk}) \times \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}''}^{t'k}, \sigma_\ell) \,, \qquad (11)$$

that imposes a spatial smoothness constraint on the flow field via adaptive spatial weighting of motion estimations from coarser scale. The combination of both potentials (8) and (11) results in the complete conditional flow field transition probability as given in (2).

We impose adaptive spatial constraints on every factor of the V-transition. The transition factors (8) and (11) allow us to *unroll* two different kinds of spatial constraints along the temporal and the scale axes while adapting the uncertainties for scale and time transition differently. This is done by splitting not only the transition in two pairwise potentials, one for the temporal- and one for the scale-transition, but also every potential in itself in two factors, one for the transition noise and the other one for an additional spatial constraint. In this way, the coupling of the potentials (8) and (11) realizes a combination of (A) scale-time prediction and (B) an integration of motion information neighboring in time, in space and in scale.

2.3 Approximate Inference

To gain a recurrent optical flow filtering we propose an approximate inference based on belief propagation [15] with factored Gaussian belief representations. The structure of the graphical model in Fig. 1 is similar to a Markov Random field. To derive a forward filter suitable for online applications we propose the following message passing scheme. Let us assume, we isolate one time slice at time t and neglect all past and future beliefs, then we would have to propagate the messages $m_{k\to k'}$ (see Fig. 1) from coarse to fine and the messages $m_{k'\to k}$ from fine to coarse to compute a posterior belief over the scale Markov chain. The two-dimensional scale-time filter (*STF*) combines this with forward passing of temporal messages $m_{t\to t'}$ and the computation of the likelihood messages $m_{Y\to v} = \ell(\mathbf{v}_x^{t'k'})$ at all scales k.

As a simplification we restrict ourselves to propagating messages only in one direction $k \to k'$ and neglect passing back the message $m_{k'\to k}$. The consequence of this is that not all the $\mathbf{V}$-nodes at time t have seen all the data $\mathbf{Y}^{1:t,1:K}$ but only all past data up to the current scale $\mathbf{Y}^{1:t,1:k}$. This increases computational efficiency and is a suitable approximation since we are only interested in the flow field on the finest scale $\mathbf{V}^{t,K}$ which is now the only node that sees all the data $\mathbf{Y}^{1:t,1:K}$. Nevertheless, future implementations will need to evaluate whether propagating also back will improve the accuracy significantly.

More precisely, the factored observation likelihood and the transition probability we introduced in (1) and (2) ensure that the forward propagated joint belief

$$P(\mathbf{V}^{t,1:K}|\mathbf{Y}^{1:t,1:K}) = \prod_{\mathbf{x}} P(\mathbf{v}_{\mathbf{x}}^{t,1:K}|\mathbf{Y}^{1:t,1:K}) \qquad (12)$$

will remain factored. In addition, we assume the belief over $\mathbf{V}^{tk}$ and $\mathbf{V}^{tk'}$ at time t to be factored which implies that also the belief over $\mathbf{V}^{t'k}$ and $\mathbf{V}^{tk'}$ factorizes.

$$\begin{aligned} &P(\mathbf{V}^{t'k}, \mathbf{V}^{tk'} \mid \mathbf{Y}^{1:t',1:k'} \setminus \mathbf{Y}^{t'k'}) = \\ =&P(\mathbf{V}^{t'k} \mid \mathbf{Y}^{1:t',1:k}) P(\mathbf{V}^{tk'} \mid \mathbf{Y}^{1:t,1:k'}) \qquad (13) \\ =&\prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{t'k}) \alpha(\mathbf{v}_{\mathbf{x}}^{tk'}) \,, \end{aligned}$$

where we used α's as the notation for forward filtered beliefs and $\setminus$ for excluding $\mathbf{Y}^{t'k'}$ from the set of measurements $\mathbf{Y}^{1:t',1:k'}$. The STF forward filter can now be defined by the computation of updated beliefs as the product of incoming messages,

$$\alpha(\mathbf{v}_x^{tk}) \propto m_{Y\to v}(\mathbf{v}_{\mathbf{x}}^{tk})\, m_{t\to t'}(\mathbf{v}_{\mathbf{x}}^{tk})\, m_{k\to k'}(\mathbf{v}_{\mathbf{x}}^{tk}) \,, \quad (14)$$

with

$$\begin{aligned} m_{t\to t'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) &= \int_{\mathbf{V}^{tk'}} \phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) \alpha(\mathbf{V}^{tk'}) \mathrm{d}\mathbf{V}^{tk'} \\ &= \sum_{\mathbf{x}'} \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{x}-\mathbf{x}', \mathbf{\Sigma}_{t,\mathbf{x}}^{tk}) \times \qquad (15) \\ &\int_{\mathbf{v}_{\mathbf{x}'}^{tk'}} \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}'}^{tk'}, \sigma_t) \alpha(\mathbf{v}_{\mathbf{x}'}^{tk'}) \mathrm{d}\mathbf{v}_{\mathbf{x}'}^{tk'} \,, \end{aligned}$$

$$\begin{aligned} m_{k\to k'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) &= \int_{\mathbf{V}^{t'k}} \phi_\ell(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) \alpha(\mathbf{V}^{t'k}) \mathrm{d}\mathbf{V}^{t'k} \\ &= \sum_{\mathbf{x}''} \mathcal{N}(\mathbf{x}''|\mathbf{x}, \mathbf{\Sigma}_{\ell,\mathbf{x}}^{tk}) \times \qquad (16) \\ &\int_{\mathbf{v}_{\mathbf{x}''}^{t'k}} \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}''}^{t'k}, \sigma_\ell) \alpha(\mathbf{v}_{\mathbf{x}''}^{t'k}) \mathrm{d}\mathbf{v}_{\mathbf{x}''}^{t'k} \,. \end{aligned}$$

For reasons of computational complexity we introduce a last approximative restriction. We want every factor of the posterior probability (14) to be Gaussian distributed

$$\begin{aligned} \alpha(\mathbf{v}_x^{tk}) &\propto m_{Y\to v}(\mathbf{v}_{\mathbf{x}}^{tk})\, m_{t\to t'}(\mathbf{v}_{\mathbf{x}}^{tk})\, m_{k\to k'}(\mathbf{v}_{\mathbf{x}}^{tk}) \\ &:\approx \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{tk}|\boldsymbol{\mu}_{\mathbf{x}}^{tk}, \mathbf{\Sigma}_{\mathbf{x}}^{tk}) \,. \qquad (17) \end{aligned}$$

We fulfill this constraint by making all single messages Gaussian distributed[1]. This already holds for the observation likelihood $m_{Y\to v}(\mathbf{v}_\mathbf{x}^{tk})$. Inserting Gaussian distributed beliefs α into the propagation equations (15, 16) leads to two different *Mixture of Gaussians* (MoG's) for the resulting messages

$$\begin{aligned} m_{t\to t'}(\mathbf{v}_\mathbf{x}^{t'k'}) &= \sum_{\mathbf{x}'} \hat{p}_{\mathbf{x}'}^{t'k'} \mathcal{N}(\mathbf{v}_\mathbf{x}^{t'k'} | \hat{\boldsymbol{\mu}}_{\mathbf{x}'}^{t'k'}, \hat{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{t'k'}) \\ &\approx \mathcal{N}(\mathbf{v}_\mathbf{x}^{t'k'} | \boldsymbol{\omega}_\mathbf{x}^{t'k'}, \boldsymbol{\Omega}_\mathbf{x}^{t'k'}) \,, \end{aligned} \tag{18}$$

with

$$\hat{p}_{\mathbf{x}'}^{t'k'} = \mathcal{N}(\mathbf{x} - \mathbf{x}' | \boldsymbol{\mu}_{\mathbf{x}'}^{tk'}, \check{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{tk'}) \,, \tag{19}$$

$$\hat{\boldsymbol{\mu}}_{\mathbf{x}'}^{t'k'} = (\sigma_t + \boldsymbol{\Sigma}_{\mathbf{x}'}^{tk'}) \check{\boldsymbol{\Lambda}}_{\mathbf{x}'}^{tk'} (\mathbf{x} - \mathbf{x}') + \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk} \check{\boldsymbol{\Lambda}}_{\mathbf{x}'}^{tk'} \boldsymbol{\mu}_{\mathbf{x}'}^{tk'} \,, \tag{20}$$

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{t'k'} = \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk} \check{\boldsymbol{\Lambda}}_{\mathbf{x}'}^{tk'} (\sigma_t + \boldsymbol{\Sigma}_{\mathbf{x}'}^{tk'}) \,, \tag{21}$$

$$\check{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{tk'} = \left[\check{\boldsymbol{\Lambda}}_{\mathbf{x}'}^{tk'} \right]^{-1} = \sigma_t + \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk} + \boldsymbol{\Sigma}_{\mathbf{x}'}^{tk'} \,,$$

and

$$\begin{aligned} m_{k\to k'}(\mathbf{v}_\mathbf{x}^{t'k'}) &= \sum_{\mathbf{x}''} \overline{p}_{\mathbf{x}''}^{t'k'} \mathcal{N}(\mathbf{v}_\mathbf{x}^{t'k'} | \boldsymbol{\mu}_{\mathbf{x}''}^{t'k}, \overline{\boldsymbol{\Sigma}}_{\mathbf{x}''}^{t'k'}) \\ &\approx \mathcal{N}(\mathbf{v}_\mathbf{x}^{t'k'} | \boldsymbol{\pi}_\mathbf{x}^{t'k'}, \boldsymbol{\Pi}_\mathbf{x}^{t'k'}) \,, \end{aligned} \tag{22}$$

with

$$\overline{p}_{\mathbf{x}''}^{t'k'} = \mathcal{N}(\mathbf{x}'' | \mathbf{x}, \boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk}) \,, \quad \overline{\boldsymbol{\Sigma}}_{\mathbf{x}''}^{t'k'} = \sigma_\ell + \boldsymbol{\Sigma}_{\mathbf{x}''}^{t'k} \,. \tag{23}$$

In order to satisfy the Gaussian constraint formulated in (17) the MoG's are collapsed into single Gaussians (18, 22) again. This is derived by minimizing the Kullback-Leibler Divergence between the given MoG's and the assumed Gaussians for the means $\boldsymbol{\omega}_\mathbf{x}^{tk}, \boldsymbol{\pi}_\mathbf{x}^{tk}$ and the covariances $\boldsymbol{\Omega}_\mathbf{x}^{tk}, \boldsymbol{\Pi}_\mathbf{x}^{tk}$ which results in closed-form solutions for these parameters. The final *predictive belief* $\alpha(\mathbf{v}_x^{tk})$ follows from the product of these Gaussians

$$\alpha(\mathbf{v}_x^{tk}) = \ell(\mathbf{v}_x^{tk}) \, \mathcal{N}(\mathbf{v}_\mathbf{x}^{tk} | \tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk}, \tilde{\boldsymbol{\Sigma}}_\mathbf{x}^{tk}) \,, \tag{24}$$

$$\tilde{\boldsymbol{\Sigma}}_\mathbf{x}^{tk} = \boldsymbol{\Pi}_\mathbf{x}^{tk} \left[\boldsymbol{\Pi}_\mathbf{x}^{tk} + \boldsymbol{\Omega}_\mathbf{x}^{tk} \right]^{-1} \boldsymbol{\Omega}_\mathbf{x}^{tk} \,, \tag{25}$$

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk} =& \boldsymbol{\Omega}_\mathbf{x}^{tk} \left[\boldsymbol{\Pi}_\mathbf{x}^{tk} + \boldsymbol{\Omega}_\mathbf{x}^{tk} \right]^{-1} \boldsymbol{\pi}_\mathbf{x}^{tk} + \\ & \boldsymbol{\Pi}_\mathbf{x}^{tk} \left[\boldsymbol{\Pi}_\mathbf{x}^{tk} + \boldsymbol{\Omega}_\mathbf{x}^{tk} \right]^{-1} \boldsymbol{\omega}_\mathbf{x}^{tk} \,. \end{aligned} \tag{26}$$

By applying the approximation steps (17, 18) and (22) we guarantee the posterior (14) to be Gaussian which allows for Kalman-filter like update equations since the observation is defined to factorize into Gaussian factors (3). The final recurrent motion estimation is given by

$$\begin{aligned} \alpha(\mathbf{v}_\mathbf{x}^{tk}) =& \mathcal{N}(\mathbf{v}_\mathbf{x}^{tk} \mid \boldsymbol{\mu}_\mathbf{x}^{tk}, \boldsymbol{\Sigma}_\mathbf{x}^{tk}) \\ =& \mathcal{N}(-\mathbf{I}_{t,\mathbf{x}}^{tk} \,|\, (\nabla \mathbf{I}_\mathbf{x}^{tk})^T \mathbf{v}_\mathbf{x}^{tk}, \boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk}) \times \\ & \mathcal{N}(\mathbf{v}_\mathbf{x}^{tk} \mid \tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk}, \tilde{\boldsymbol{\Sigma}}_\mathbf{x}^{tk}) \,, \end{aligned} \tag{27, 28}$$

$$\boldsymbol{\Sigma}_\mathbf{x}^{tk} = \left[\tilde{\boldsymbol{\Lambda}}_\mathbf{x}^{tk} + \nabla \mathbf{I}_\mathbf{x}^{tk} \boldsymbol{\Lambda}_{\ell,\mathbf{x}}^{tk} (\nabla \mathbf{I}_\mathbf{x}^{tk})^T \right]^{-1} \,, \tag{29}$$

$$\boldsymbol{\mu}_\mathbf{x}^{tk} = \tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk} - \boldsymbol{\Sigma}_\mathbf{x}^{tk} \nabla \mathbf{I}_\mathbf{x}^{tk} \boldsymbol{\Lambda}_{\ell,\mathbf{x}}^{tk} \tilde{\mathbf{I}}_{t,\mathbf{x}}^{tk} \,. \tag{30}$$

For reasons explained in [11] the innovations process is approximated as the following

$$\tilde{\mathbf{I}}_{t,\mathbf{x}}^{tk} \approx \partial / \partial t \mathcal{T} \left(\mathbf{I}_\mathbf{x}^{tk}, \tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk} \right) \,, \tag{31}$$

with $\mathcal{T}$ applying a backward warp plus bilinear interpolation on the image $\mathbf{I}_\mathbf{x}^{tk}$ using the predicted velocities $\tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk}$ from (26). What we gain is a general probabilistic scale-time filter (STF) which is, in comparison to existent filtering approaches [7], [11], [13], not a Kalman Filter realization but a Dynamic Bayesian Network. If we have access to a batch of data (or a recent window of data) and do not focus on online-oriented pure forward filtering we can compute smoothed posteriors $\gamma(\mathbf{v}_\mathbf{x}^{tk}) := P(\mathbf{v}_\mathbf{x}^{tk} | \mathbf{Y}^{1:T,1:k})$. Therefore, we follow a Two-Filter realization for optical flow smoothing as proposed in [14].

3 Adaptivity Information

Now that we have set up probabilistic filtering equations (30, 29) for recurrent optical flow computation that constrain the estimation based on the extended Lucas-Kanade assumption that the movement within a multidimensional $(\mathbf{x},\ k,\ t)$ neighborhood is constant, we continue to specify the neighborhood relations. As defined in section 2 we want the integration of neighboring velocity estimates to be adaptable in scale k, time t and location $\mathbf{x}$. Therefore, the corresponding covariances $\boldsymbol{\Sigma}_{I,\mathbf{x}}^{tk},\ \boldsymbol{\Sigma}_{t,\mathbf{x}}^{t'k},\ \boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk'}$ of the different Gaussian kernels are adapted dependent on the local structural information of the underlying intensity patches $\mathbf{I}_\mathbf{x}^{tk}$ within the neighborhood.

We assume that neighbors along the orientation of the local structure are more likely to influence the velocity of the center pixel than neighbors that are located beside the orientation. For this reason, we increase the spatial uncertainty for the location of the center pixel along the orientation of the structure by increasing the uncertainty of the covariance matrices $\boldsymbol{\Sigma}_{I,\mathbf{x}}^{tk},\ \boldsymbol{\Sigma}_{t,\mathbf{x}}^{t'k},\ \boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk'}$ aligned with the orientation. On the other hand, we reduce the spatial uncertainty orthogonal to the orientation to strengthen the assumption that we are more certain that the position of the pixel is somewhere

[1] A more accurate technique (following assumed density filtering) would be to first compute the new belief α exactly as a MoGs and then collapse it to a single Gaussian. However, this would mean extra costs. Future research will need to investigate the tradeoff between computational cost and accuracy for different collapsing methods.

along (as compared to orthogonally to) the structural contour. To obtain the orientation θ of the local structure we use the local Structure Tensor $\mathbf{H}^{tk}$ averaged with a Gaussian $\mathbf{G}_{\iota}$

$$\mathbf{H}^{tk} = \mathbf{G}_{\iota} * \begin{pmatrix} (\mathbf{I}_{\chi}^{tk})^2 & \mathbf{I}_{\chi}^{tk}\mathbf{I}_{y}^{tk} \\ \mathbf{I}_{\chi}^{tk}\mathbf{I}_{y}^{tk} & (\mathbf{I}_{y}^{tk})^2 \end{pmatrix} , \tag{32}$$

perform an eigendecomposition for every $\mathbf{H}_{\mathbf{x}}^{tk}$ and get the eigenvalues $\lambda_{1,\mathbf{x}}^{tk}, \lambda_{2,\mathbf{x}}^{tk}$ and the eigenvectors $(\cos\theta_{\mathbf{x}}^{tk},\ \sin\theta_{\mathbf{x}}^{tk})^T$, $(-\sin\theta_{\mathbf{x}}^{tk},\ \cos\theta_{\mathbf{x}}^{tk})^T$ for every scale k, time t and position $\mathbf{x}$. Similar to an edge enhancing diffusion tensor $\mathbf{D}_{\mathbf{x}}^{tk}$ [4] we calculate oriented covariances as follows

$$\mathbf{\Sigma}_{i,\mathbf{x}}^{tk} := \mathbf{D}_{\mathbf{x}}^{tk} = f(\theta_{\mathbf{x}}^{tk}, g(\lambda_{1,\mathbf{x}}^{tk}), \kappa_{i,2}) , \tag{33}$$

$$g(\lambda_{1,\mathbf{x}}^{tk}) = \frac{\kappa_{i,2}}{1 + (\lambda_{1,\mathbf{x}}^{tk}/\kappa_{i,1})^4} , \quad i \in \{\ell, \kappa, \iota\} . \tag{34}$$

This leads to covariance matrices aligned with the underlying intensity structure and prefers to group velocity information along *motion contours* and not across *motion discontinuities* because in most cases it is true that motion contours overlap with intensity contours.

4 Evaluation

We present some performance results based on the Middlebury benchmark for optical flow evaluation [1] to argue the applicability of our probabilistic scale-time filter STF. In all the experiments the parameters are chosen fixed[2]. The uncertainty on the flow field $\mathbf{\Sigma}_v$ is chosen to be a diagonal matrix with entries s_v. The kernel sizes for the spatial derivatives are chosen to be 5×5 and for the temporal derivative 3×3 with filter coefficients as proposed in [10]. The size of the adaptive spatial Gaussian filters is chosen $(2^k+1) \times (2^k+1)$, dependent on the scale k and the Gaussian smoothing kernel of the Structure Tensor $\mathbf{G}_{\iota}$ has a diagonal covariance matrix with standard deviation of 2.5 at a kernel size of 5×5.

In table 1 we report some error statistics for the *cloudless Yosemite* sequence (see Fig. 2) using the same error measures as proposed by [1]. As can be seen, we are able to keep up with the accuracy of recent optical flow methods. The overall performance (see also Fig. 3 a)) of STF with an average angular error (AAE) of $1.52°$ outperforms high accuracy optical flow techniques, like the popular CLG method of Bruhn et al. [5]. Looking at the performance only at the motion discontinuities shown in Fig. 3 b) the AAE is with $3.07°$ still quite low. This argues for the spatiotemporal adaptation of the uncertainties as described in Sec. 3 which allows to keep the motion discontinuities pronounced. Other methods like the Black and Anandan approach [3] that introduce robust statistics to reduce the errors at motion boundaries because of motion outliers seem to be less effective. Nevertheless, in untextured regions the CLG method that includes the global optical flow constraint of Horn and Schunk [8] gets better results with an AAE of $1.46°$ compared to $1.53°$ using STF which has no additional global constraint in the observation likelihood measurement. The explanation for that is as follows: The STF method is a probabilistic recurrent filter that takes into account neighboring measurements for optical flow predictions to the next time frame. Therefore, it realizes a filling-in process over time via a predictive prior but without a further global smoothness constraint on the measurement. The more data with constant image flow is processed over time the larger is the *propagation range* into untextured regions. Thus, for a small number of filter steps, such an incorporation of local smoothness propagating along image location in time seems to be less effective than a direct incorporation like obtained via the Horn and Schunk constraint.

Another interesting result is shown in Fig. 4. Here, the performance for different graduations of the STF filter is shown. (A) and (B) are the time lapses for a belief propagation filter *only along scale* which neglects the temporal messages $m_{t \to t'}$. In case (A) the spatial filters are not adapted which is equivalent to $\kappa_{i,1} \to \infty$ and in case (B) they are adapted like explained in Sec. 3. In case (C) the forward filter results are shown without uncertainty adaptation to the

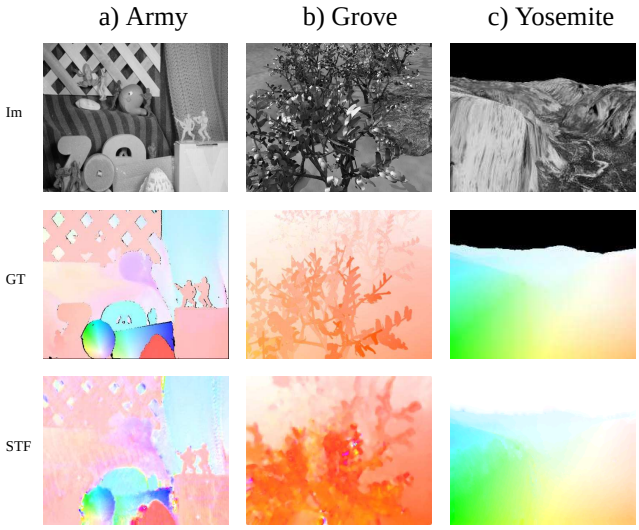

Figure 2. Qualitative comparison between ground truth (GT) and the STF method for different benchmark sequences (Im) a) Army b) Grove and c) Yosemite from the Middlebury database.

[2] $\kappa_{\ell,1} = 10^3$, $\kappa_{\kappa,1} = 2 \cdot 10^3$, $\kappa_{\iota,1} = 2 \cdot 10^2$, $\kappa_{\ell,2} = 2.25$, $\kappa_{\kappa,2} = 1.5$, $\kappa_{\iota,2} = 60$, $\sigma_{\kappa} = 0.1$, $\sigma_{\iota} = 10^{-12}$, $s_v = s_{\iota} = 10^{-3}$.

technique	frame number	angular error all	angular error disc	angular error untext
Lucas & Kanade	2	6.41°	7.02°	10.8°
LP Registration	2	4.51°	5.48°	3.95°
Horn & Schunk	2	4.01°	5.41°	1.95°
Dynamic MRF	2	3.63°	5.29°	4.62°
Black & Anandan	2	2.61°	4.44°	2.15°
2D CLG	2	1.76°	3.14°	1.46°
STF	8	1.52°	3.07°	1.53°

Table 1. Results of state-of-the-art methods for the *cloudless Yosemite* sequence and our results for a batch of 8 frames applying adaptive Two-Filter inference STF with 6 filter steps in time t and 3 along scale k.

local structure and in (D) the forward filter results with uncertainty adaptation can be seen. Both, the local adaptation and the spatiotemporal prediction via $m_{t\to t'}$ improves the performance.

Figure 5 clarifies that motion estimation is refined as more data is aquired while propagating beliefs over scale as well as forward (A) and backward (B) in time. Of course, the best results are gained if future as well as past data is taken into account to estimate the flow at current time, like in the two-filter approach (C). Taking a closer look at the time lapse of the AAE for the adaptive purely scale-propagation filter (see Fig. 4 (B)) compared to the time lapse of the AAE for the adaptive two-filter (see Fig. 5 (C)) it turns out that the $mean$ and the standard deviation std of the AAE for all frames are higher for the time-isolated scale filter $mean = 3.64$, $std = 0.29$ compared to the time-dependent two-filter $mean = 2.18$, $std = 0.25$. This means, beside quite large variations of the AAE over time because of changing errors at motion boundaries reflected in changing observation likelihood measurements our STF reduces the variance of the AAE over time. Such kind of improvement is not possible for methods that are based on an *isolated* batch of images which do not allow for estimation changes because of new arriving evidence.

5 Conclusion

We have presented a new recurrent filter for optical flow estimation (STF) which incrementally improves the estimation accuracy based on scale-time predictions and adapts to the structure of the observed scene. The performance of our filter is superior compared to standard Lucas-Kanade and comparable to combined local-global methods like CLG. The consequent exploitation of the local Lucas-Kanade constraint by STF fills in untextured regions to a certain amount

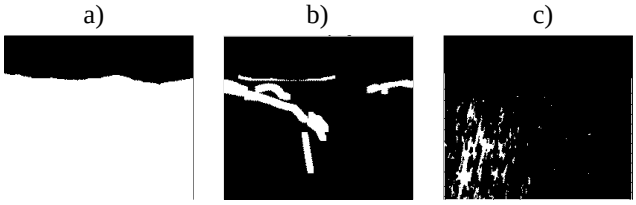

Figure 3. The white color marks a) all the pixels b) only the discontinuities (disc) or c) only the untextured (untext) regions which are chosen for the results given in table 1.

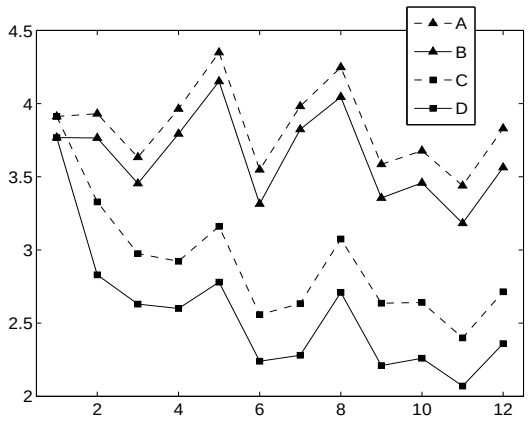

Figure 4. Time lapse of the average angular error for (A) the nonadaptive STF method over scale (B) the STF method over scale (C) the nonadaptive STF method and (D) the complete adaptive forward STF method.

and keeps being very accurate at motion discontinuities. The main advantage of the STF method lies in the online applicability and the adaptation to movement changes similar to object tracking approaches. In particular, the STF filter realises a probabilistic tracking of the whole dense optical flow field.

References

[1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R.Szeliski. A database and evaluation methodology for optical flow. In *ICCV*, 2007.

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

[3] M. Black. Recursive non-linear estimation of discontinuous flow fields. In *ECCV*, pages 138–145, 1994.

[4] T. Brox and J. Weickert. Nonlinear matrix diffusion for optic flow estimation. In *Pattern Recognition, Lecture Notes in Computer Science*, pages 446–453, 2002.

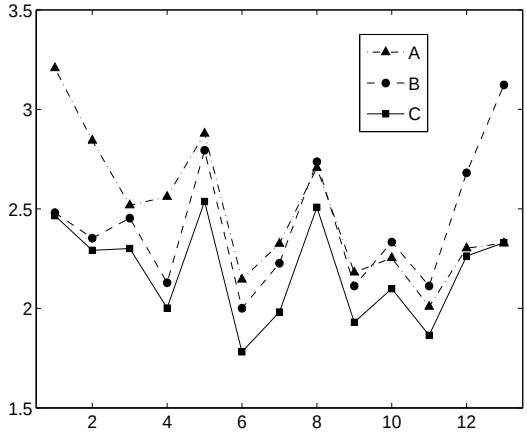

Figure 5. Time lapse of the average angular error for (A) the adaptive forward filter (B) the adaptive backward filter and (C) the adaptive two-filter inference.

[5] 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.

[6] A. Doshi and A. Bors. Optical flow diffusion with robustified kernels. In *Computer Analysis of Images and Patterns*, pages 222–230, 2005.

[7] M. Elad and A. Feuer. Recursive optical flow estimation-adaptive filtering approach. *Journal of Visual Communication and image representation*, 9:119–138, 1998.

[8] B. K. P. Horn and B. G. Schunk. Determining optic flow. *Artificial Intelligence*, 17:185–204, 1981.

[9] B. D. Lukas and T. Kanade. An iterative image-registration technique with an application to stereo vision. In *IJCAI*, pages 674–679, Vancouver, Canada, 1981.

[10] H. Scharr. Optimal filters for extended optical flow. In *First International Workshop on Complex Motion*, pages 185–204, 2004.

[11] E. Simoncelli. *Handbook of Computer Vision and Applications*, chapter Bayesian Multi-Scale Differential Optical Flow, pages 397–421. Academic Press, 1999.

[12] E. Simoncelli, E. Adelson, and D. Heeger. Probability distributions of optical flow. In *CVPR*, pages 310–315, 1991.

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

[14] V. Willert, M. Toussaint, J. Eggert, and E. Körner. Uncertainty optimization for robust dynamic optical flow estimation. In *ICMLA*, pages 450–457, 2007.

[15] J. Yedidia, W. Freeman, and Y. Weiss. *Exploring Artificial Intelligence in the New Millennium*, chapter Understanding Belief Propagation and Its Generalizations, pages 239–236. Morgan Kaufmann, 2003.