A Stochastic Dynamical System for Optical Flow Estimation

Volker Willert
TU Darmstadt, Landgraf-Georg-Str. 4
64283 Darmstadt, Germany
vwillert@rtr.tu-darmstadt.de

Julian Eggert
Honda Research Institute Europe, Carl-Legien-Str. 30
63073 Offenbach/Main, Germany
julian.eggert@honda-ri.de

Abstract

So far, the research on optical flow has mainly concentrated on motion estimations using the observation of a small number of temporal consecutive frames of an image sequence. The dynamics of the flow field evolution is mostly neglected. Our main concern is to stress that visual motion is a dynamic feature of an image input stream and the more visual data has been observed the more precise and detailed we can estimate and predict the motion contained in the visual data. In this paper, we present a probabilistic dynamical system that is suitable to recurrently infer visual motion. The assumed flow dynamics fuses spatial smoothness constraints and smoothness constraints along time and scale. We propose a certain class of transition probability functions which satisfy a probability mixture model and allow for an efficient approximate inference based on Belief Propagation. We arrive at a compact and general algorithm for optical flow filtering and realize one instance using factored Gaussian belief representations.

1. Introduction

Despite many years of progress, visual motion processing inspires a lot of interest in researchers involved in understanding the principles of visual perception. Basic aspects such as measuring optical flow investigating local smoothness constraints have been widely studied. But what is most striking about motion processing is its temporal dynamics. This is obvious, because the environment perceived by a visual observer like a moving video camera or the human eye is highly dynamic. Moving objects enter and leave the field of view and also temporal changes of the camera position and view point induce visual motion. Hence, suitable assumptions about the dynamics of the visual scene and about the correlations between local flow vectors are beneficial for the estimation of visual motion. The main problem for optical flow estimation is to resolve the ambiguities in the measurement process. It is an ill-posed problem because of noisy visual data and insufficient models that are assumed for the relation between visual data and the optical flow. To be able to model the ambiguities several authors propose to introduce uncertainty, for example [13]. For this purpose, the velocity of an image location and the images of a sequence are understood as statistical signals. It allows to explicitly consider noise in the measurement processes and to incorporate prior knowledge about the state to be estimated. Usually, such priors incorporate smoothness assumptions on the flow field. There are *three basic smoothness constraints* often found in the literature: Smoothness between image velocities defined 1) along different points in image location, 2) different points in time, or 3) different image scales. The most established method to reduce ambiguities is the spatial smoothness constraint [8, 9]. This is often accounted for by smoothness constraints for neighboring velocities assuming that all pixels within some spatial neighborhood move similarly. Further improvements are made using multiscale approaches [11, 4]. This is desirable, e.g., for being able to represent both large and small velocities at coarse and fine resolutions with a reasonable effort. Another important aspect of motion estimation is temporal continuity. This has motivated several researchers [3, 7, 6] to recursively estimate the optical flow over time including a prediction model that defines some temporal relation between pixel movements. Recently, Markov Random Fields regained great popularity to impose spatial smoothness priors on motion measurements [10]. On the contrary side, the idea to treat optical flow estimation as a dynamical system, like in Kalman-Filter approaches, received few consideration, lately. Here, we propose to fuse both ideas - spatial smoothness and smoothness along time and scale - into one common predictive prior model. This allows the formulation of a probabilistic dynamical system to infer visual motion via spatiotemporal belief propagation. The main contribution is the proposal of a certain class of transition probability functions which satisfy a probability mixture model and allow for temporal prediction along with spatial smoothing. For this class of transitions we derive a computationally tractable optical flow filter. We show results in terms of accuracy, robustness and long range applicability

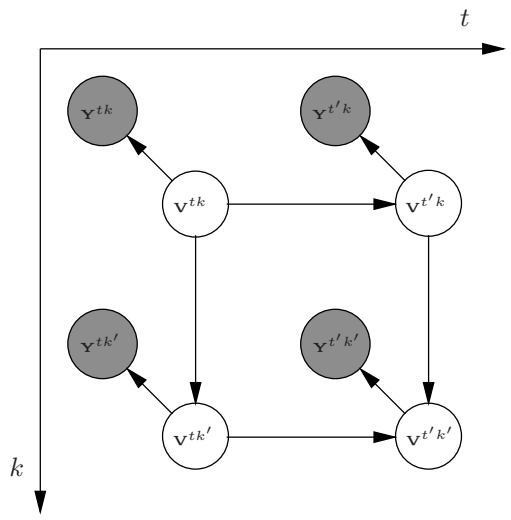

Figure 1. A probabilistic directed graphical model for visual motion estimation. Here, $t' = t + 1$ denotes next future timestep and $k' = k + 1$ denotes next finer scale. Observable nodes representing image data are shaded gray, hidden nodes representing optical flows are colored white.

for a filter realization with a factored Gaussian belief representation.

2. A dynamical system for motion estimation

Having all the different possibilities of motion disambiguation in mind, the question is how to unify them in one general framework for motion estimation. In this section, we derive a probabilistic solution which is computationally tractable.

Dynamic Bayesian Network A Dynamic Bayesian Network (DBN) is a directed graphical model of a dynamic stochastic process. Here, we propose such a network as depicted in figure 1 to model the dynamics of visual motion. The structure of the graphical model in figure 1 is similar to a Markov Random field. The difference is that the edges are directed. It tightly couples several *Markov chains along time* that are defined for each scale k via *Markov chains along scale* defined at each time step t. Note that the DBN forces an independency structure. The probability that a node is in one of its states depends directly only on the states of its parents [14]. 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 $t \in \mathcal{T}$ at K spatial resolution scales $k \in \mathcal{K}$ with $t' = t + \Delta t$ and $k' = k + \Delta k$ being the next time step and the next finer scale, respectively. Without loss of generalization we define the time intervals $\Delta t = 1$ and scale intervals $\Delta k = 1$ to be unity. Here, the observable $\mathbf{Y}^{tk}$ comprises image data of several frames within a time interval around t at the same scale k. For example, $\mathbf{Y}^{tk} = (\mathbf{I}^{tk}, \mathbf{I}^{t'k})$ has to be at least a pair of images with both images are defined over the same image range X^k at the same scale k but at consecutive points in time t and t'. Each image $\mathbf{I}^{tk}$ consists of image intensities $I^{tk}_{\mathbf{x}}$ at each image position $\mathbf{x} \in \mathcal{X}^k$. Similarly, the hidden state $\mathbf{V}^{tk}$ is a flow field at time slice t and scale k defined over the image range X^k with velocity vectors $\mathbf{v}^{tk}_{\mathbf{x}}$ at each image position $\mathbf{x}$.

Generative model The probabilistic generative model is precisely defined by the following probabilities and factorization assumptions: First, an *initial prior* for the flow field at time $t = 1$ and scale $k = 1$

$$P(\mathbf{V}^{11}) = \prod_{\mathbf{x}} P(\mathbf{v}^{11}_{\mathbf{x}}) \,, \tag{1}$$

defining some preference for the speed and direction of the velocities in the flow field. Often this is chosen to be a product of zero mean Gaussian distributions to prefer slow and smooth velocities [13]. Second, the specification of the *observation likelihood* for the images $\mathbf{Y}^{tk}$ given the flow $\mathbf{V}^{tk}$ for all times $t \in \mathcal{T}$ and scales $k \in \mathcal{K}$

$$P(\mathbf{Y}^{tk}|\mathbf{V}^{tk}) = \prod_{\mathbf{x}} \ell(\mathbf{Y}^{tk}, \mathbf{v}^{tk}_{\mathbf{x}}) \,. \tag{2}$$

This factorisation assumption is somewhat unusual because we do not assume the image observation to factorize in pixel observations but assume the observation likelihood to factorize in the velocities only. And third, the specification of the *transition probabilities* for the flow fields $\mathbf{V}^{t'k'}$ at the new timestep t' at finer scale k' 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 first time slice $t = 1$ and the coarsest scale $k = 1$ the transitions are conditioned only on one flow field $\mathbf{V}^{1k}$ or $\mathbf{V}^{t1}$.

$$P(\mathbf{V}^{1k'}|\mathbf{V}^{1k}) \propto \prod_{\mathbf{x}} \phi_k(\mathbf{v}^{1k'}_{\mathbf{x}}, \mathbf{V}^{1k}) \,, \tag{3}$$

$$P(\mathbf{V}^{t'1}|\mathbf{V}^{t1}) \propto \prod_{\mathbf{x}} \phi_t(\mathbf{v}^{t'1}_{\mathbf{x}}, \mathbf{V}^{t1}) \,, \tag{4}$$

$$P(\mathbf{V}^{t'k'}|\mathbf{V}^{t'k}, \mathbf{V}^{tk'}) \propto \prod_{\mathbf{x}} \phi_k(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{V}^{t'k}) \times \phi_t(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{V}^{tk'}) \,. \tag{5}$$

These equations explicitly express that the probability distribution for each flow field factorizes into independent distributions for each velocity vector. Nevertheless, although each velocity vector is not dependent on velocity vectors from the flow field at the same time and scale it heavily depends on all the velocity vectors from the flow fields at coarser scale and past time. Further on, the conditional dependence $P(\mathbf{v}^{t'k'}_{\mathbf{x}}|\mathbf{V}^{t'k}, \mathbf{V}^{tk'})$ can be split in two pairwise potentials ϕ_k, ϕ_t. This will allow us to maintain only factored beliefs during inference, which makes the approach computationally practicable.

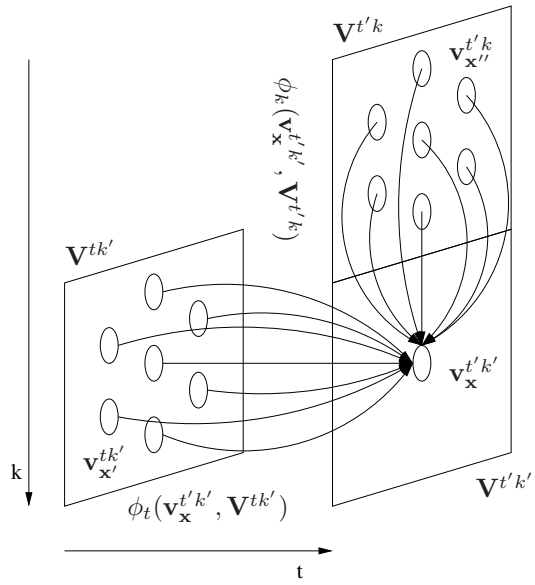

Figure 2. Pairwise potentials of the scale-time transition.

A general class of flow field transitions To further specify the generative model we have to define the formulas for the prior (1), the observation likelihood (2), and the transitions (5). A concrete example is given in the next section 3. For the flow field transitions in equation (5) we propose a certain class of transition probability functions which satisfy a probability mixture model. Equation (5) consists of two pairwise potentials. The first potential $\phi_t(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{V}^{tk'})$ assumes that the flow field at every spatial scale k transforms from $t \rightarrow t'$ according to itself. The second potential $\phi_k(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{V}^{t'k})$ realizes a refinement from coarser to finer scale $k \rightarrow k'$ at every time t'. A sketch of the information flow is shown in figure 2. To motivate the temporal transition factor $\phi_t(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{V}^{tk'})$ we assume that the direction and speed of a flow vector $\mathbf{v}^{t'k'}_{\mathbf{x}}$ at position $\mathbf{x}$ at time t' is functionally related to a previous flow vector $\mathbf{v}^{tk'}_{\mathbf{x}'}$ at some corresponding position $\mathbf{x}'$ at time t,

$$\mathbf{v}^{t'k'}_{\mathbf{x}} \sim f_t(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{v}^{tk'}_{\mathbf{x}'}; \theta_t) \,. \tag{6}$$

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

$$\mathbf{x}' \sim f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}^{t'k'}_{\mathbf{x}}; \theta_{xt}) \,. \tag{7}$$

In principle f_t and f_{xt} can be any arbitrary functions that define the relation between spatiotemporally neighboring velocities. The free parameters θ_t and θ_{xt} allow for adaptation of the temporal and spatial relations, respectively. Note that here we use $\mathbf{v}^{t'k'}_{\mathbf{x}}$ to retrieve the previous corresponding point $\mathbf{x}'$. This is a suitable approximation as long as the similarity $\mathbf{v}^{t'k'}_{\mathbf{x}} \approx \mathbf{v}^{tk'}_{\mathbf{x}'}$ is not heavily violated. Combining both factors (6) and (7) and integrating $\mathbf{x}'$ leads to the first pairwise potential

$$\phi_t(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{V}^{tk'}) = \sum_{\mathbf{x}'} f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}^{t'k'}_{\mathbf{x}}; \theta_{xt}) f_t(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{v}^{tk'}_{\mathbf{x}'}; \theta_t) \,. \tag{8}$$

Equivalent to (6) for the scale transition factor $\phi_k(\mathbf{v}^{1k'}_{\mathbf{x}}, \mathbf{V}^{1k})$ we assume that the origin of a flow vector $\mathbf{v}^{t'k'}_{\mathbf{x}}$ at position $\mathbf{x}$ at finer scale k' corresponds to a flow vector $\mathbf{v}^{t'k}_{\mathbf{x}''}$ from coarser scale k at some corresponding position $\mathbf{x}''$,

$$\mathbf{v}^{t'k'}_{\mathbf{x}} \sim f_k(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{v}^{t'k}_{\mathbf{x}''}; \theta_k) \,. \tag{9}$$

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)

$$\mathbf{x}'' \sim f_{xk}(\mathbf{x}'', \mathbf{x}; \theta_{xk}) \,. \tag{10}$$

The considerations for the scale transition are analogous to the ones for the temporal transition. Again, combining both factors (9) and (10) and integrating $\mathbf{x}''$ we get the second pairwise potential

$$\phi_k(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{V}^{t'k}) = \sum_{\mathbf{x}''} f_{xk}(\mathbf{x}'', \mathbf{x}; \theta_{xk}) f_k(\mathbf{v}^{t'k'}_{\mathbf{x}}, \mathbf{v}^{t'k}_{\mathbf{x}''}; \theta_k) \,, \tag{11}$$

that imposes a spatial smoothness constraint on the flow field via 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 (5). The transition factors (8) and (11) allow us to *unroll* two different kinds of spatial constraints along the temporal and the scale axes (see figure 2) while adapting the free parameters 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.

Inference The overall data likelihood $P(\mathbf{Y}^{1:T,1:K}, \mathbf{V}^{1:T,1:K})$ is assumed to factorize as defined by the directed graph shown in figure 1

$$P(\mathbf{Y}^{1:T,1:K}, \mathbf{V}^{1:T,1:K}) = \prod_{t=1}^{T} \prod_{k=1}^{K} P(\mathbf{Y}^{tk}|\mathbf{V}^{tk}) P(\mathbf{V}^{11}) \times$$
$$\prod_{t=1}^{T-1} P(\mathbf{V}^{t'1}|\mathbf{V}^{t1}) \prod_{k=1}^{K-1} P(\mathbf{V}^{1k'}|\mathbf{V}^{1k}) P(\mathbf{V}^{t'k'}|\mathbf{V}^{t'k}, \mathbf{V}^{tk'}) \,. \tag{12}$$

What we are usually interested in is the probability for some flow field given all the data aquired so far. For the forward-backward filtered (offline) case [2] where all future and past data of a sequence is accessable this would be the probability $P(\mathbf{V}^{tk}|\mathbf{Y}^{1:T,1:K})$. For the forward filtered (online) case [2] where only the past data is accessable the probability $P(\mathbf{V}^{tk}|\mathbf{Y}^{1:t,1:K})$ would be interesting. To infer these probabilities, Bayes' rule and marginalization has to be applied. For the offline case this reads

$$P(\mathbf{V}^{tk}|\mathbf{Y}^{1:T,1:K}) = \sum_{\substack{\mathbf{V}^{1:T,1:K} \\ \setminus \mathbf{V}^{tk}}} \frac{P(\mathbf{Y}^{1:T,1:K}, \mathbf{V}^{1:T,1:K})}{P(\mathbf{Y}^{1:T,1:K})}\,, \quad (13)$$

where we used $\setminus$ as the notation for excluding $\mathbf{V}^{tk}$ from the set of optical flow fields $\mathbf{V}^{1:T,1:K}$. The online case neglects future observations and simplifies to

$$P(\mathbf{V}^{tk}|\mathbf{Y}^{1:t,1:K}) = \sum_{\substack{\mathbf{V}^{1:t,1:K} \\ \setminus \mathbf{V}^{tk}}} \frac{P(\mathbf{Y}^{1:t,1:K}, \mathbf{V}^{1:t,1:K})}{P(\mathbf{Y}^{1:t,1:K})}\,. \quad (14)$$

Either for the online or offline case, the direct computation of the marginals using equation (13) or (14) would take exponential time [14]. One solution to this problem is *Belief Propagation* which is an efficient approximate inference algorithm especially applicable if the graph has a lot of loops and many hidden nodes like it is the case for our graphical model for dynamic motion estimation (see figure 1).

Approximate inference using Belief Propagation Here, we propose an approximate inference algorithm based on Belief Propagation (BP) and restrict ourselves to the online case (14) since its extension to the offline case is straightforward [2]. The marginal probabilities that are now computed only approximately are called *beliefs* and here we use α's as the notation for forward filtered beliefs

$$\alpha(\mathbf{v}_{\mathbf{x}}^{tk}) \approx P(\mathbf{v}_{\mathbf{x}}^{tk}|\mathbf{Y}^{1:t,1:K})\,. \quad (15)$$

Let us start with the inference of the flow field at first time slice $t = 1$ and coarsest scale $k = 1$ just having access to the observable $\mathbf{Y}^{11}$. Applying Bayes' rule we get

$$\alpha(\mathbf{v}_{\mathbf{x}}^{11}) = P(\mathbf{v}_{\mathbf{x}}^{11}|\mathbf{Y}^{11}) = \frac{\ell(\mathbf{Y}^{11}, \mathbf{v}_{\mathbf{x}}^{11}) P(\mathbf{v}_{\mathbf{x}}^{11})}{P(\mathbf{Y}^{11})}\,. \quad (16)$$

This is the initial belief that has to be propagated along time and scale. To derive an approximate forward filter suitable for online applications we propose the following message passing scheme [14] that realizes a recurrent update of the beliefs. 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 $m_{k\to k'}$ from coarse to fine and the messages $m_{k'\to k}$ from fine to coarse to compute a belief over the scale Markov chain. Similarly, if we isolate one scale k for all time slices and neglect all coarser and finer beliefs, then we would have to propagate the messages $m_{t\to t'}$ from the past to the future and the messages $m_{t'\to t}$ from the future to the past to compute a belief over the temporal Markov chain. For the realization of a forward scale-time filter, we combine the forward passing of temporal messages $m_{t\to t'}$ and the computation of the likelihood messages $m_{Y\to v} = \ell(\mathbf{Y}^{t'k'}, \mathbf{v}_x^{t'k'})$ 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 reduces computational costs but the flow field on the finest scale $\mathbf{V}^{t,K}$ is now the only node that sees all the data $\mathbf{Y}^{1:t,1:K}$. Nevertheless, we also tested passing back the messages $m_{k'\to k}$ which only slightly improved the accuracy but increased computational costs.

More precisely, the factored observation likelihood and the transition probability we introduced in (2) and (5) ensure that the forward propagated joint belief

$$P(\mathbf{V}^{t,1:K}|\mathbf{Y}^{1:t,1:K}) = \prod_{\mathbf{x}} P(\mathbf{v}_{\mathbf{x}}^{t,1:K}|\mathbf{Y}^{1:t,1:K}) \quad (17)$$

will remain factored. Similar to BP in a Markov Random Field, we assume independency for all neighboring nodes in the Markov blanket. This means the belief over $\mathbf{V}^{tk}$ and $\mathbf{V}^{tk'}$ at time t is assumed to be factored which implies that also the belief over $\mathbf{V}^{t'k}$ and $\mathbf{V}^{tk'}$ factorizes.

$$P(\mathbf{V}^{t'k}, \mathbf{V}^{tk'}|\mathbf{Y}^{1:t',1:k'} \setminus \mathbf{Y}^{t'k'}) = \quad (18)$$

$$P(\mathbf{V}^{t'k}|\mathbf{Y}^{1:t',1:k}) P(\mathbf{V}^{tk'}|\mathbf{Y}^{1:t,1:k'}) = \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{t'k}) \alpha(\mathbf{v}_{\mathbf{x}}^{tk'})\,,$$

where we used $\setminus$ as the notation for excluding $\mathbf{Y}^{t'k'}$ from the set of measurements $\mathbf{Y}^{1:t',1:k'}$. The two-dimensional forward filter propagates the belief over $\mathbf{V}^{t'k}$ and $\mathbf{V}^{tk'}$ from (18) via multiplying with the scale-time transition (5) and marginalizing over $\mathbf{V}^{t'k}$ and $\mathbf{V}^{tk'}$. The result is multiplied with the new observation likelihood (2) and normalized by $P(\mathbf{Y}^{t'k'})$ to get the updated belief as shown in equations (19)-(22).

As can be seen, the complete scale-time forward filter can now be defined by the computation of updated beliefs α as the product of incoming messages,

$$\alpha(\mathbf{v}_x^{tk}) \propto m_{Y\to v}(\mathbf{v}_{\mathbf{x}}^{tk})\; m_{k\to k'}(\mathbf{v}_{\mathbf{x}}^{tk})\; m_{t\to t'}(\mathbf{v}_{\mathbf{x}}^{tk})\,. \quad (23)$$

Inserting the proposed class of temporal transitions (8) into (22) leads to the derivation of the temporal message equation

$$m_{t\to t'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) = \sum_{\mathbf{V}^{tk'}} \phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{tk'})\,, \quad (24)$$

$$P(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{Y}^{1:t',1:k'}) = \frac{1}{P(\mathbf{Y}^{t'k'})} \quad \ell(\mathbf{Y}^{t'k'}, \mathbf{v}_{\mathbf{x}}^{t'k'}) \sum_{\mathbf{V}^{t'k}} \sum_{\mathbf{V}^{tk'}} P(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{V}^{t'k}, \mathbf{V}^{tk'}) P(\mathbf{V}^{t'k}, \mathbf{V}^{tk'}|\mathbf{Y}^{1:t',1:k'} \setminus \mathbf{Y}^{t'k'}), \quad (19)$$

$$\alpha(\mathbf{v}_{\mathbf{x}}^{t'k'}) \propto m_{Y\to v}(\mathbf{v}_{\mathbf{x}}^{t'k'}) \sum_{\mathbf{V}^{t'k}} \sum_{\mathbf{V}^{tk'}} \overbrace{\phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) \phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'})}^{} \overbrace{\prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{t'k}) \alpha(\mathbf{v}_{\mathbf{x}}^{tk'})}^{}, \quad (20)$$

$$\propto m_{Y\to v}(\mathbf{v}_{\mathbf{x}}^{t'k'}) \underbrace{\sum_{\mathbf{V}^{t'k}} \phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{t'k})}_{} \underbrace{\sum_{\mathbf{V}^{tk'}} \phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{tk'})}_{}, \quad (21)$$

$$\propto m_{Y\to v}(\mathbf{v}_{\mathbf{x}}^{t'k'}) \quad \times \quad m_{k\to k'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) \quad \times \quad m_{t\to t'}(\mathbf{v}_{\mathbf{x}}^{t'k'}). \quad (22)$$

$$= \sum_{\mathbf{V}^{tk'}} \sum_{\mathbf{x}'} f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}; \theta_{xt}) \times f_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{tk'}; \theta_t) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{tk'}), \quad (25)$$

$$= \sum_{\mathbf{v}_{\mathbf{x}}^{tk'}} \sum_{\mathbf{x}'} f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}; \theta_{xt}) \times f_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{tk'}; \theta_t) \alpha(\mathbf{v}_{\mathbf{x}}^{tk'}) \underbrace{\sum_{\substack{\mathbf{V}^{tk'} \\ \setminus \mathbf{v}_{\mathbf{x}'}^{tk'}}} \prod_{\mathbf{z}\neq\mathbf{x}'} \alpha(\mathbf{v}_{\mathbf{z}}^{tk'})}_{1}, \quad (26)$$

$$= \sum_{\mathbf{x}'} f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}; \theta_{xt}) \times \sum_{\mathbf{v}_{\mathbf{x}}^{tk'}} f_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{tk'}; \theta_t) \alpha(\mathbf{v}_{\mathbf{x}'}^{tk'}). \quad (27)$$

Note that the summation $\sum_{\mathbf{V}^{tk'}}$ is summing over all possible flow fields, i.e. $\sum_{\mathbf{V}^{tk'}}$ represents X^k summations $\sum_{\mathbf{v}_{1,1}^{tk'}} \sum_{\mathbf{v}_{1,2}^{tk'}} \sum_{\mathbf{v}_{2,1}^{tk'}} \cdots$ over each local flow field vector. We separated these into a summation $\sum_{\mathbf{v}_{\mathbf{x}'}^{tk'}}$ over the flow field vector at $\mathbf{x}'$ and a summation $\sum_{\mathbf{V}^{tk'}\setminus\mathbf{v}_{\mathbf{x}'}^{tk'}}$ over all other flow field vectors at $\mathbf{x} \neq \mathbf{x}'$. Then, we use the equivalence $\sum_{\mathbf{V}^{tk'}\setminus\mathbf{v}_{\mathbf{x}'}^{tk'}} \prod_{\mathbf{z}\neq\mathbf{x}'} \alpha(\mathbf{v}_{\mathbf{z}}^{tk'}) = \prod_{\mathbf{z}\neq\mathbf{x}'} \sum_{\mathbf{v}_{\mathbf{z}}^{tk'}} \alpha(\mathbf{v}_{\mathbf{z}}^{tk'}) = 1$. Similarly, we arrive at the scale message if we insert the scale transition (11) into (22)

$$m_{k\to k'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) \propto \sum_{\mathbf{x}'} f_{xk}(\mathbf{x}', \mathbf{x}; \theta_{xk}) \times \sum_{\mathbf{v}_{\mathbf{x}}^{t'k}} f_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{t'k}; \theta_k) \alpha(\mathbf{v}_{\mathbf{x}'}^{t'k}). \quad (28)$$

Finally, the three equations (23), (27), and (28) define a tightly coupled scale-time forward filter for visual motion estimation. It realizes a probabilistic recurrent estimation of a set of flow fields $\mathbf{V}^{t,1:K}$ with different resolutions k swept along the time dimension t. The following pseudo-code shows the compact form of the derived scale-time filter

Algorithm 1 Pseudo-code for the scale-time filter

```
Initialize the priors α(v_x^{0,1:K})
for t' = 1 to T do
  for k' = 1 to K do
    for x = 1 to X^{k'} do
      Compute the messages
      m_{Y→v}(v_x^{t'k'})
      m_{t→t'}(v_x^{t'k'}) according to equation (27)
      m_{k→k'}(v_x^{t'k'}) according to equation (28)
      Update the beliefs
      α(v_x^{t'k'}) according to equation (23)
    end for
  end for
end for
```

suitable for an algorithmic implementation. What remains to be done, is the specification of the observation likelihood (2) and the potentials of the transition probability (8) and (11).

3. The Gaussian filter realisation

Now, we define the observation likelihood and the transitions in such a way that we are left with a purely Gaussian belief representation. This results in a filter similar to an extended Kalman Filter only propagating means and covariances along scale and time.

Gaussian observation likelihood We follow a similar argumentation as Simoncelli *et al.* [12] to obtain the $\ell(\mathbf{v}_{\mathbf{x}}^{tk})$-factors (2) 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.

$$\ell(\mathbf{v}_{\mathbf{x}}^{tk}) = \mathcal{N}(-\mathbf{I}_{t,\mathbf{x}}^{tk}|(\nabla\mathbf{I}_{\mathbf{x}}^{tk})^T\mathbf{v}_{\mathbf{x}}^{tk}, \boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk}) , \quad (29)$$

$$\boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk} = \begin{pmatrix} \ddots & \dots & \mathbf{0} \\ \vdots & \sigma_{\ell,\mathbf{x}\mathbf{x}'}^{tk} & \vdots \\ \mathbf{0} & \dots & \ddots \end{pmatrix} , \quad (30)$$

$$\sigma_{\ell,\mathbf{x}\mathbf{x}'}^{tk} = \frac{(\nabla I_{\mathbf{x}'}^{tk})^T\mathbf{S}_v\nabla I_{\mathbf{x}'}^{tk} + s_t}{f_\ell(\mathbf{x}', \mathbf{x}, t, k)} . \quad (31)$$

In notation (29), the patches can be regarded as vectors and the covariance matrix $\boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk}$ 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. The likelihood formulation defines multivariate Gaussian distributions for vectors that describe image patches centered around image locations. Allowing for uncertainties $\boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk}$ that are adaptive in location $\mathbf{x}$, scale k and time t we are able to tune the local motion measurements dynamically e.g. dependent on the underlying structure of the intensity patterns.

Mixture of Gaussians transition For the temporal contraint (6) we now chose a Gaussian

$$\mathbf{v}_{\mathbf{x}}^{t'k'} \sim \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}'}^{tk'}, \sigma_t) , \quad (32)$$

which says that the change in time of the flow field is white with undirectional transition noise between $\mathbf{V}^{tk'}$ and $\mathbf{V}^{t'k'}$. For the spatial interaction (7) an inhomogeneous anisotropic Gaussian is assumed

$$\mathbf{x}' \sim \mathcal{N}(\mathbf{x}'|\mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}, \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk}) . \quad (33)$$

to be able to steer the orientation and to adapt the strength of the uncertainty in spatial identification $\boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk}$ between corresponding positions in time. Combining both factors (32) and (33) and integrating $\mathbf{x}'$ we get a *Mixture of Gaussians* (MoG) as the first pairwise potential (8)

$$\phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) = \sum_{\mathbf{x}'} \mathcal{N}(\mathbf{x}'|\mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}, \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk}) \times \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}'}^{tk'}, \sigma_t) , \quad (34)$$

with the Gaussian spatial coherence constraint being the mixing coefficients. Equivalent to (32) for the scale transition factor (11) we chose a Gaussian

$$\mathbf{v}_{\mathbf{x}}^{t'k'} \sim \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}''}^{t'k}, \sigma_k) , \quad (35)$$

assuming white transition noise σ_k. The influence of neighboring velocity states from coarser scale is also modelled as an adaptive Gaussian kernel similar to (33)

$$\mathbf{x}'' \sim \mathcal{N}(\mathbf{x}''|\mathbf{x}, \boldsymbol{\Sigma}_{k,\mathbf{x}}^{tk}) . \quad (36)$$

Again, combining both factors (35) and (36) and integrating $\mathbf{x}''$ we get a MoG as the second pairwise potential

$$\phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) = \sum_{\mathbf{x}''} \mathcal{N}(\mathbf{x}''|\mathbf{x}, \boldsymbol{\Sigma}_{k,\mathbf{x}}^{tk}) \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}''}^{t'k}, \sigma_k) , \quad (37)$$

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 (5).

Approximate inference To arrive at a Gaussian belief we introduce a last approximative restriction. We want every factor of the posterior probability (23) to be Gaussian distributed

$$\alpha(\mathbf{v}_x^{tk}) \propto m_{Y\to v}(\mathbf{v}_{\mathbf{x}}^{tk})\, m_{t\to t'}(\mathbf{v}_{\mathbf{x}}^{tk})\, m_{k\to k'}(\mathbf{v}_{\mathbf{x}}^{tk}) \approx \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{tk}|\boldsymbol{\mu}_{\mathbf{x}}^{tk}, \boldsymbol{\Sigma}_{\mathbf{x}}^{tk}) . \quad (38)$$

We fulfill this constraint by making all single messages Gaussian distributed. This already holds for the observation likelihood $m_{Y\to v}(\mathbf{v}_{\mathbf{x}}^{tk})$. A more accurate technique (following assumed density filtering) would be to first compute the new belief α exactly as a MoG and then collapse it to a single Gaussian. However, this would mean extra costs. Here, we do not investigate the tradeoff between computational cost and accuracy for different collapsing methods. Inserting Gaussian distributed beliefs α into the propagation equations (27), (28) leads to two different MoGs for the resulting messages

$$m_{t\to t'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) = \sum_{\mathbf{x}'} \hat{p}_{\mathbf{x}'}^{t'k'} \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\hat{\boldsymbol{\mu}}_{\mathbf{x}'}^{t'k'}, \hat{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{t'k'}) \approx \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\boldsymbol{\omega}_{\mathbf{x}}^{t'k'}, \boldsymbol{\Omega}_{\mathbf{x}}^{t'k'}) , \quad (39)$$

with

$$\hat{p}_{\mathbf{x}'}^{t'k'} = \mathcal{N}(\mathbf{x} - \mathbf{x}'|\boldsymbol{\mu}_{\mathbf{x}'}^{tk'}, \check{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{tk'}) , \quad (40)$$

$$\hat{\boldsymbol{\mu}}_{\mathbf{x}'}^{t'k'} = (\sigma_t + \boldsymbol{\Sigma}_{\mathbf{x}'}^{tk'})\check{\boldsymbol{\Lambda}}_{\mathbf{x}'}^{tk'}(\mathbf{x} - \mathbf{x}') + \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk}\check{\boldsymbol{\Lambda}}_{\mathbf{x}'}^{tk'}\boldsymbol{\mu}_{\mathbf{x}'}^{tk'} , \quad (41)$$

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{t'k'} = \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk}\check{\boldsymbol{\Lambda}}_{\mathbf{x}'}^{tk'}(\sigma_t + \boldsymbol{\Sigma}_{\mathbf{x}'}^{tk'}) , \quad (42)$$

$$\check{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{tk'} = \left[\check{\boldsymbol{\Lambda}}_{\mathbf{x}'}^{tk'}\right]^{-1} = \sigma_t + \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk} + \boldsymbol{\Sigma}_{\mathbf{x}'}^{tk'} , \quad (43)$$

and

$$m_{k\to k'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) = \sum_{\mathbf{x}''} \overline{p}_{\mathbf{x}''}^{t'k'} \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\boldsymbol{\mu}_{\mathbf{x}''}^{t'k}, \overline{\boldsymbol{\Sigma}}_{\mathbf{x}''}^{t'k'}) \approx \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\boldsymbol{\pi}_{\mathbf{x}}^{t'k'}, \boldsymbol{\Pi}_{\mathbf{x}}^{t'k'}) , \quad (44)$$

with

$$\overline{p}_{\mathbf{x}''}^{t'k'} = \mathcal{N}(\mathbf{x}''|\mathbf{x}, \boldsymbol{\Sigma}_{k,\mathbf{x}}^{tk}) , \quad \overline{\boldsymbol{\Sigma}}_{\mathbf{x}''}^{t'k'} = \sigma_k + \boldsymbol{\Sigma}_{\mathbf{x}''}^{t'k} . \quad (45)$$

In order to satisfy the Gaussian constraint formulated in (38) the MoG's are collapsed into single Gaussians (39),

(44) 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

$$\alpha(\mathbf{v}_x^{tk}) = \ell(\mathbf{v}_x^{tk})\, \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{tk} | \tilde{\boldsymbol{\mu}}_{\mathbf{x}}^{tk}, \tilde{\boldsymbol{\Sigma}}_{\mathbf{x}}^{tk}) \,, \quad (46)$$

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{x}}^{tk} = \boldsymbol{\Pi}_{\mathbf{x}}^{tk} \left[\boldsymbol{\Pi}_{\mathbf{x}}^{tk} + \boldsymbol{\Omega}_{\mathbf{x}}^{tk}\right]^{-1} \boldsymbol{\Omega}_{\mathbf{x}}^{tk} \,, \quad (47)$$

$$\tilde{\boldsymbol{\mu}}_{\mathbf{x}}^{tk} = \boldsymbol{\Omega}_{\mathbf{x}}^{tk} \left[\boldsymbol{\Pi}_{\mathbf{x}}^{tk} + \boldsymbol{\Omega}_{\mathbf{x}}^{tk}\right]^{-1} \boldsymbol{\pi}_{\mathbf{x}}^{tk} + \boldsymbol{\Pi}_{\mathbf{x}}^{tk} \left[\boldsymbol{\Pi}_{\mathbf{x}}^{tk} + \boldsymbol{\Omega}_{\mathbf{x}}^{tk}\right]^{-1} \boldsymbol{\omega}_{\mathbf{x}}^{tk} \,. \quad (48)$$

By applying the approximation steps (38), (39) and (44) we guarantee the posterior (23) to be Gaussian which allows for Kalman-filter like update equations since the observation is defined to factorize into Gaussian factors (29). The final recurrent motion estimation is given by

$$\alpha(\mathbf{v}_{\mathbf{x}}^{tk}) = \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{tk} | \boldsymbol{\mu}_{\mathbf{x}}^{tk}, \boldsymbol{\Sigma}_{\mathbf{x}}^{tk}) \quad (49)$$

$$= \mathcal{N}(-\mathbf{I}_{t,\mathbf{x}}^{tk} | (\nabla \mathbf{I}_{\mathbf{x}}^{tk})^T \mathbf{v}_{\mathbf{x}}^{tk}, \boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk}) \times \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{tk} | \tilde{\boldsymbol{\mu}}_{\mathbf{x}}^{tk}, \tilde{\boldsymbol{\Sigma}}_{\mathbf{x}}^{tk}) \,, \quad (50)$$

$$\boldsymbol{\Sigma}_{\mathbf{x}}^{tk} = \left[\tilde{\boldsymbol{\Lambda}}_{\mathbf{x}}^{tk} + \nabla \mathbf{I}_{\mathbf{x}}^{tk} \boldsymbol{\Lambda}_{\ell,\mathbf{x}}^{tk} (\nabla \mathbf{I}_{\mathbf{x}}^{tk})^T\right]^{-1} \,, \quad (51)$$

$$\boldsymbol{\mu}_{\mathbf{x}}^{tk} = \tilde{\boldsymbol{\mu}}_{\mathbf{x}}^{tk} - \boldsymbol{\Sigma}_{\mathbf{x}}^{tk} \nabla \mathbf{I}_{\mathbf{x}}^{tk} \boldsymbol{\Lambda}_{\ell,\mathbf{x}}^{tk} \tilde{\mathbf{I}}_{t,\mathbf{x}}^{tk} \,. \quad (52)$$

The innovations process can be approximated as the following

$$\tilde{\mathbf{I}}_{t,\mathbf{x}}^{tk} \approx \partial / \partial t \mathcal{T} \left(\mathbf{I}_{\mathbf{x}}^{tk}, \tilde{\boldsymbol{\mu}}_{\mathbf{x}}^{tk}\right) \,, \quad (53)$$

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 (48). We end up with a Gaussian scale-time filter which is related to an extended Kalman Filter since the result of the nonlinear transitions is linearized after each message pass with the collpase of each MoG to a single Gaussian.

4. Evaluation

We present some performance results based on the Yosemite benchmark and real world sequences to argue the applicability of our probabilistic scale-time filter. The covariances $\boldsymbol{\Sigma}_{I,\mathbf{x}}^{tk}$, $\boldsymbol{\Sigma}_{t,\mathbf{x}}^{t'k}$, $\boldsymbol{\Sigma}_{k,\mathbf{x}}^{tk'}$ of the different Gaussian kernels are chosen to be equivalent to an edge enhancing diffusion tensor $\mathbf{D}_{\mathbf{x}}^{tk}$ at different scales and times. All other parameters are fixed during runtime. In table 1 we report some error statistics for the *cloudless Yosemite* sequence using the same error measures as proposed by [1] and added the results of Simoncelli [11]. As can be seen, we are able to keep up with the accuracy of recent optical flow methods. The overall performance with an average angular error (AAE) of 1.52° keeps up with high accuracy optical flow techniques, like the CLG method of Bruhn et al. [5]. Looking at the

technique	frame number	angular error all	angular error disc	angular error untext
Lukas & 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°
Simoncelli	2	3.81°	–	–
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°
our approach	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 6 filter steps in time t and 3 along scale k.

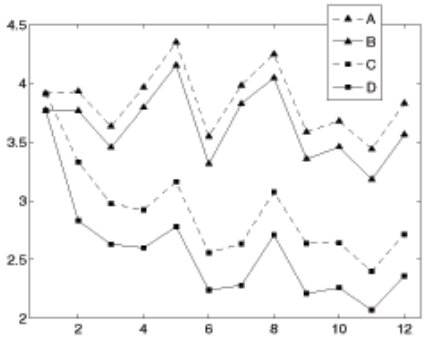

Figure 3. Time lapse of the average angular error for (A) the nonadaptive method over scale (B) the adaptive method over scale (C) the nonadaptive method and (D) the complete adaptive forward method.

performance only at the motion discontinuities like given in [1] the AAE is with 3.07° quite low. 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 our approach which has no additional global constraint in the observation likelihood measurement. The explanation for that is as follows: The probabilistic filter 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 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 result is shown in Fig. 3. Here, the performance for different graduations of the 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

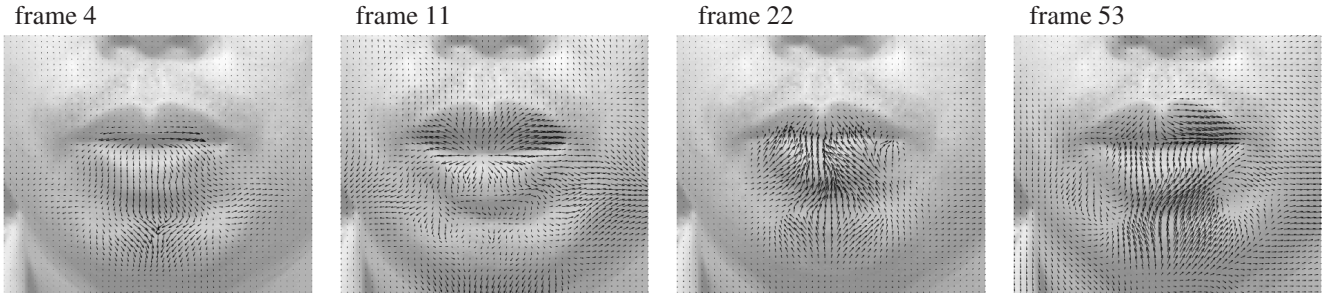

Figure 4. Motion estimation examples computed with the proposed scale-time filter along an image sequence comprising 62 frames.

(A) the spatial filters are not adapted and in case (B) they are adapted dependent on the diffusion tensor. In case (C) the forward filter results are shown without uncertainty adaptation to the 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.

Taking a closer look at the time lapse of the AAE for the adaptive purely scale-propagation filter (B) compared to the time lapse of the AAE for the adaptive forward filter (D) 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 forward filter $mean = 2.18$, $std = 0.25$. This means, beside large variations of the AAE over time because of changing errors at motion boundaries reflected in changing observation likelihood measurements our filter 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.

A real world example is given in figure 4. Estimation results for the *moving mouth* sequence are shown that contains 62 frames and shows mouth movements of spoken words. The scale-time filter adapts to the motion changes over time and produces smooth estimation results without outliers.

5. Conclusion

One key capability of a motion estimation system is the adaptability to temporal and structural changes. This is due to the fact that variations happen in observer motion, in observed object motion, and in spatial resolution as moving objects come closer to the observer and detailed texture becomes more apparent. We presented a new recurrent filter for optical flow estimation which incrementally improves the estimation accuracy based on scale-time predictions and adapts to the structure of the observed scene. The main advantage of the method lies in the online applicability and the adaptation to movement changes similar to object tracking approaches. In particular, the proposed 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] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, 2006.

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

[4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, pages 25–36, 2004.

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

[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, 1981.

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

[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] 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.

[14] 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.