Dynamic Visual Motion Estimation

Volker Willert, Julian Eggert

2010

Chapter 1

Dynamic Visual Motion Estimation

1.1 Introduction

Visual motion is the projection of scene movements on a visual sensor. It is a rich source of information for the analysis of a visual scene. Especially for dynamic vision systems the estimation of visual motion is important because it allows to deduce the motion of objects as well as the self-motion of the system relative to the environment. Therefore, visual motion serves as a basic information for navigation and exploration tasks, like obstacle avoidance, object tracking or visual scene decomposition into static and moving parts.

Despite many years of progress, visual motion processing continues to puzzle the mind of researchers involved in understanding the principles of visual perception. Basic aspects such as measuring motions of spatially local entities 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 video camera or the human eye is highly dynamic. Moving objects enter and leave the field of view and also change the way they move, e.g. change the direction or speed. Hence, suitable assumptions about the dynamics of the visual scene and about the correlations between local moving entities are beneficial for the estimation of the scene motion as a whole.

Probabilistic machine learning techniques have become very popular for early vision problems like binocular depth and optical flow computation. The reason for that popularity is because of the possibility to explicitly consider inherent uncertainties in the measurement processes and to incorporate prior knowledge about the state to be estimated. Along this line of argumentation, we present a general approach to visual motion estimation based on a probabilistic generative model that allows to infer visual motion from visual data. We start with a definition of visual motion and point out the basic problems that come along with visual motion estimation. Then, we summaries common ideas that can be found in different state-of-the-art optical flow estimation techniques and stress the need for taking uncertainty into account. Based on the ideas of already existing models we introduce a general Bayesian framework for dynamic optical flow estimation that comprises several different aspects for solving the optical flow estimation problem into one common approach.

So far, the research on optical flow has mainly concentrated on motion estimations using the observation of two frames of an image sequence isolated in time. 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 this visual data. Therefore both motion prediction and observation integration should be explicitly modeled in an inference procedure for optical flow estimation. Recently - because of the development of efficient
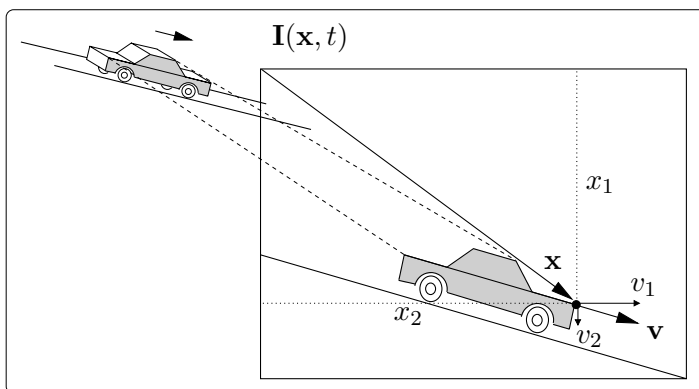
Figure 1.1: How a moving object, e.g. a car, is projected onto a surface.

loopy belief propagation algorithms - Markov Random Fields regained great popularity to impose smoothness priors on motion measurements. On the contrary side, a way to impose consistent priors that has not been paid much attention lately is to treat the motion estimation as a dynamical system, like in Kalman-Filter approaches, to propagate motion information along time or scale assuming smoothness along the time or scale dimension.

In this chapter, 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 combined with additional factorization assumptions and approximate inference techniques it is possible to get a computationally tractable probabilistic optical flow filter. To show the capability, the benefits, and the drawbacks of the framework, we derive two realizations: one for continuous Gaussian and one for discrete grid-based observation likelihoods as well as Mixture of Gaussians and Mixture of Student's-t-distributions transitions.

1.2 Problems of optical flow computation

1.2.1 Visual motion

If we move or if objects move through our field of view then the projection of the environment onto the retina of our eyes changes. The projections over time form patterns of spatiotemporal brightness changes that encode the movement of the projections called the visual motion [14]. This matter of fact has motivated researchers from diverse areas like psychology of perception, visual neuroscience or computer science to understand the principles and model the mechanisms behind visual motion estimation.

More technically spoken, image sequences captured from a camera stream are mappings from the three-dimensional world onto a two-dimensional light-sensing surface - usually a digital camera-chip - over time. The image brightness $I(\mathbf{x}, t)$ at each spatial position $\mathbf{x} = (x_1\ x_2)^T$ at a particular time t (or over some interval of time) is a measurement how much light fell on the surface. When an object in the world moves relative to this projection surface, the two-dimensional projection of that object moves within the image sequence. The movement of the projected position of each point in the world

refers to a velocity vector $\mathbf{v}(\mathbf{x}, t) = (v_1 \ v_2)^T$ and the set of all these vectors is called the motion field [24]. Figure 1.1 shows a sketch of the relations just explained.

1.2.2 Brightness constancy assumption

As mentioned beforehand, the motion field can be estimated using spatiotemporal brightness patterns. In principle, the quality of any estimation problem depends on the unambiguity of the relation between the observed data and the state to be infered [15]. In our case this is the relation between the observed spatiotemporal brightness and the motion field. The basic approach to visual motion estimation assumes constant brightness during the time brightness is measured [11]. This assumption implies that changes in brightness can only be caused by translational movements of the projections onto the image surface with constant velocity. Using the brightness constancy assumption the relation between the brightness $I(\mathbf{x}, t)$ of an image pixel at position $\mathbf{x}$ at time t and the corresponding velocity vector $\mathbf{v}(\mathbf{x}, t)$ during the time interval Δt reads

$$I(\mathbf{x} + \mathbf{v}(\mathbf{x}, t)\Delta t, t + \Delta t) = I(\mathbf{x}, t) \,. \tag{1.1}$$

If the time interval Δt is sufficiently small then the left side of equation (1.1) can be approximated via a first order Taylor series [23] as follows

$$\nabla_{\mathbf{x}} I(\mathbf{x}, t)\mathbf{v}(\mathbf{x}, t) + \frac{\partial I(\mathbf{x}, t)}{\partial t} = 0 \,. \tag{1.2}$$

Here, $\nabla_{\mathbf{x}} I(\mathbf{x}, t) = (\partial I(\mathbf{x}, t)/\partial x_1, \partial I(\mathbf{x}, t)/\partial x_2)$ denotes the spatial brightness gradient and $\partial I(\mathbf{x}, t)/\partial t$ the temporal derivative of the spatiotemporal brightness pattern. Both equations (1.1) and (1.2) are ill-posed problems because they cannot be solved for the velocity unambiguously. Further more, the displacement $\mathbf{v}\Delta t$ is only an approximation to the true visual flow, called the optical flow [4], because the brightness could also change due to changes in the lightning conditions. Nevertheless, if there is (i) constant illumination during the time interval Δt (ii) perpendicular projection of the reflected light onto the image surface and (iii) purely translational motion of the object parallel to the image surface, then optical flow is equivalent to the true motion field [4]. There are further approaches to optical flow computation like the phase-based approach which describes flow in fourier domain. A detailed description of all the different methods as well as exhaustive comparisons can be found in [3].

1.2.3 Motion ambiguity

Besides incomplete models (1.1, 1.2) there are a series of fundamental problems concerning motion estimation. The movement of an isolated pixel cannot be estimated without considering its neighboring pixels. Therefore, only using the brightness constancy assumption (1.1) or the continuity equation (1.2) is insufficient. The structure of the local brightness pattern composed by the neighboring pixels directly affects the uncertainty of the motion estimate of the center pixel. The velocity $\mathbf{v}(\mathbf{x}, t)$ is only unique if the brightness within the brightness pattern varies along the two spatial dimensions. That means, the brightness gradient $\nabla I(\mathbf{x}, t)$ must not be zero for any dimension $\mathbf{x} = (x_1 \ x_2)^T$. If this condition is not satisfied it is not possible to find an unambiguous correspondence between pixels (brightness patterns) from temporal consecutive image frames, also called the aperture or correspondence problem [13].

In figure 1.2 four types of motion uncertainties because of an ambiguity in the correspondence of brightness patterns are shown. Here, a car drives down a road and the motion of four different parts A-D of the car should be estimated. Although all parts move with the same velocity the uncertainties
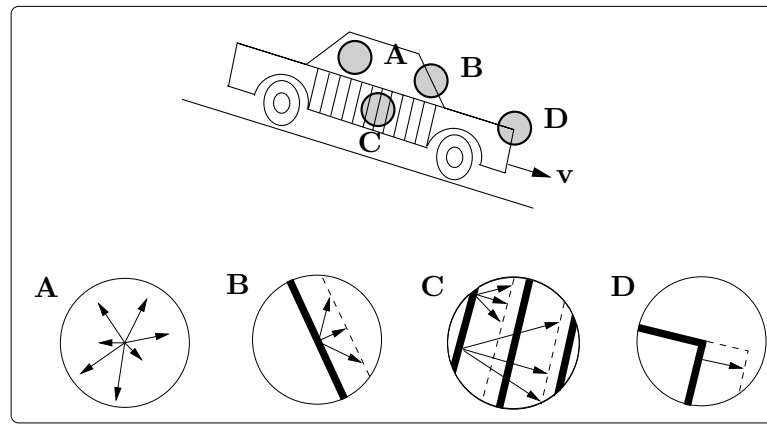
Figure 1.2: Different types of motion uncertainties because of ambiguous correspondence of brightness patterns.
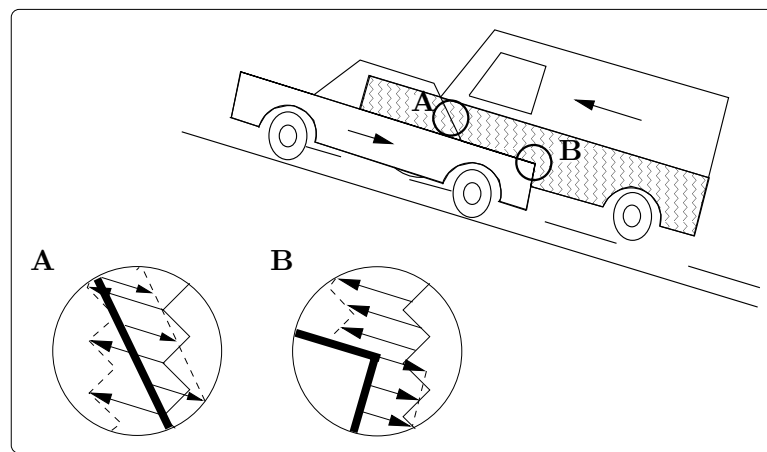

Figure 1.3: Motion ambiguities because of incoherently moving image parts and occlusion at motion discontinuities.

of the estimates differ. Part A exhibits no structure in brightness at all. Therefore, the brightness is constant over both spatial position and time. In this case, the brightness places no constraint on the velocity [24]. The local brightness of parts B and C varies only in one direction. In these cases, only the velocity component that is perpendicular to the edge direction is constrained. Along the edges the brightness does not change and thus the velocity component parallel to the edge direction cannot be estimated. In part D, the local brightness varies along two dimensions, in which case the optical flow vector is uniquely constrained and the velocity estimate gets unambiguous.

Another problem that forces ambiguities are pixel occlusions at motion boundaries. This is depicted in figure 1.3. Part A and B show two situations with brightness patterns overlapping while moving. This leads to occlusion of neighboring pixels that have been seen beforehand and newly appearing pixels that have been occluded in the past. The brightness pattern within a neighborhood of pixels along such motion boundaries are not stable over time and therefore temporal correspondences cannot be found unambiguously.

1.3   Motion disambiguation

To be able to model all these ambiguities some authors [36, 22] propose to introduce uncertainty to the model that describes the relation between image intensities and pixel velocities. For this purpose, the velocity of an image location and the images of a sequence are understood as statistical signals

$$I(\mathbf{x} + (\mathbf{v}(\mathbf{x},t) + \eta_{\mathbf{v}})\Delta t, t + \Delta t) = I(\mathbf{x},t) + \eta_I \,, \tag{1.3}$$

with $\eta_{\mathbf{v}}$ defining additive noise on the velocity and η_I additive noise on the image intensity. More general, a function F_I

$$F_I\Big(I\big(\mathbf{x} + (\mathbf{v}(\mathbf{x},t) + \eta_{\mathbf{v}})\Delta t, t + \Delta t\big), I\big(\mathbf{x},t\big)\Big) = \eta_I \,, \tag{1.4}$$

defines the relation between temporal consecutive images and the flow field. This implies probabilities for the existence of image intensities and velocities namely conditional probability density functions (pdf)

$$P\big(I(\mathbf{x}, t + \Delta t)|I(\mathbf{x},t), \mathbf{v}(\mathbf{x},t)\big) \tag{1.5}$$

that represent the uncertain relation between image observations $I(\mathbf{x}, t)$ conditioned on the hidden states $\mathbf{v}(\mathbf{x}, t)$. Using Bayes' rule and some prior knowledge about the distribution of the velocity $P(\mathbf{v}(\mathbf{x}, t))$ and the intensities $P(I(\mathbf{x}, t))$ the conditional pdf for the velocity conditioned on the image data can be infered

$$
\begin{aligned}
P\big(I(\mathbf{x}, t + \Delta t), I(\mathbf{x},t), \mathbf{v}(\mathbf{x},t)\big) &= P\big(I(\mathbf{x}, t + \Delta t)|I(\mathbf{x},t), \mathbf{v}(\mathbf{x},t)\big) \times P\big(I(\mathbf{x},t)\big) \times P\big(\mathbf{v}(\mathbf{x},t)\big) \\
&= P\big(\mathbf{v}(\mathbf{x},t)|I(\mathbf{x}, t + \Delta t), I(\mathbf{x},t)\big) \times P\big(I(\mathbf{x}, t + \Delta t)\big) \times P\big(I(\mathbf{x},t)\big) \,, \\
P\big(\mathbf{v}(\mathbf{x},t)|I(\mathbf{x}, t + \Delta t), I(\mathbf{x},t)\big) &= \frac{P\big(I(\mathbf{x}, t + \Delta t)|I(\mathbf{x},t), \mathbf{v}(\mathbf{x},t)\big) \times P\big(\mathbf{v}(\mathbf{x},t)\big)}{P\big(I(\mathbf{x}, t + \Delta t)\big)} \,.
\end{aligned}
\tag{1.6}
$$

So far, the velocities $\mathbf{v}(\mathbf{x}, t)$ and the image intensities $I(\mathbf{x}, t)$ have been assumed to be independent for different positions $\mathbf{x}$ and times t. This is already a strong approximation. In the next section this approximation is relaxed towards more spatiotemporal dependence. The expectation is that pdfs are able to tackle the addressed ambiguity problems related to motion processing. As can be seen in [23, 29, 35] specific information about the mentioned problems can, in principle, be extracted from the shape of the pdfs.

During the last ten years velocity distributions have been suggested and discussed by several authors [22, 25, 33, 30]. Additionally, a lot of different mainly deterministic approaches have been developed for visual motion estimation [4, 2]. All these approaches are based on common constraints on the flow field that are suitable for disambiguation and improvement of the estimates. The constraints F define correlations between motion estimations at different points in image location $\mathbf{x}$, different points in time t or different image scales k. Assuming independence between the dimensions $\mathbf{x}$, t, and k the constraints can be formulated in a probabilistic way as

$$F_x\big(\mathbf{v}(\mathbf{x},t,k), \{\mathbf{v}(\mathbf{x}',t,k)\}_{\mathbf{x}'}\big) = \eta_x \,, \tag{1.7}$$

$$F_t\big(\mathbf{v}(\mathbf{x},t,k), \{\mathbf{v}(\mathbf{x},t',k)\}_{t'}\big) = \eta_t \,, \tag{1.8}$$

$$F_k\big(\mathbf{v}(\mathbf{x},t,k), \{\mathbf{v}(\mathbf{x},t,k')\}_{k'}\big) = \eta_k \,. \tag{1.9}$$

Here, $\{\mathbf{v}(\mathbf{x}',t,k)\}_{\mathbf{x}'}$ denotes the set of spatial neighbors to pixel velocity $\mathbf{v}(\mathbf{x},t,k)$. The constraint is defined by the function F_x and η_x considers additive noise on the constraining model. For the set of
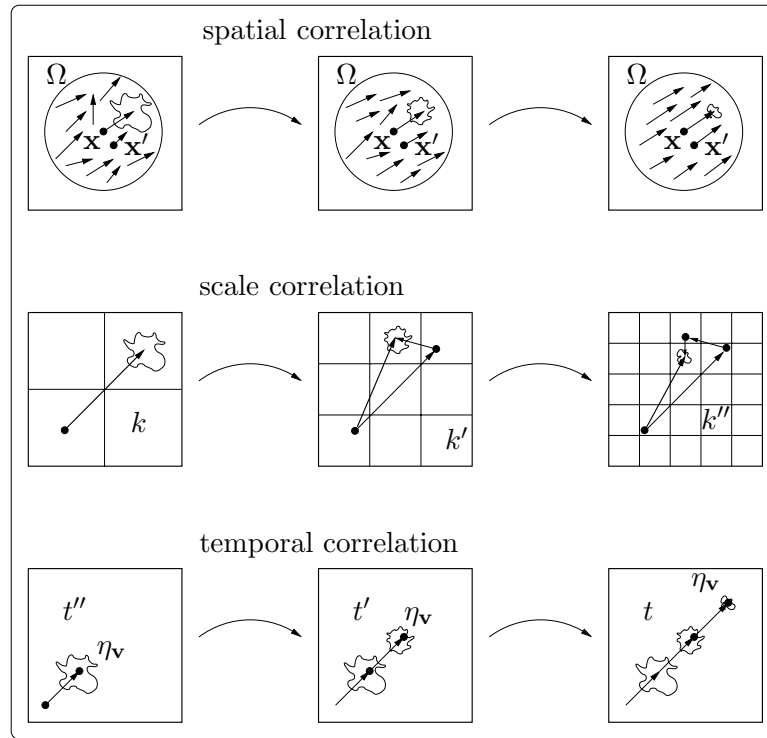
Figure 1.4: Three basic principles to disambiguate visual motion using constraints along the space, time, or scale dimension.

temporal t' and scale k' neighbors the notations are analogous. The most general case would be to assume total dependence between all dimensions

$$F\big(\mathbf{v}(\mathbf{x},t,k),\{\mathbf{v}(\mathbf{x}',t,k)\}_{\mathbf{x}'},\{\mathbf{v}(\mathbf{x},t',k)\}_{t'},\{\mathbf{v}(\mathbf{x},t,k')\}_{k'}\big)=\eta\,. \tag{1.10}$$

Unfortunately, considering full dependence together with a probabilistic treatment leads to combinatorial explosion and is in most cases computationally intractable.

A sketch of the three different types of independent disambiguation along the three dimensions is given in figure 1.4. The most established method to reduce ambiguities is the integration of motion information over space (see Fig. 1.4 first row). That means, interactions between neighboring velocities or even higher order derivations are considered [11, 4, 17, 1]. This is often accounted for by smoothness constraints for neighboring velocities assuming that all pixels within the neighborhood Ω move similarly.

Further improvements are made using multiscale approaches (see Fig. 1.4 second row). This is desirable, e.g., for being able to represent both large and small velocities at coarse and fine resolutions with a reasonable effort. It is usually done in such a way that the larger velocities at coarser scale are calculated first, then a warped version of the image is calculated using these large, coarsely sampled velocities, and afterwards the residual velocities at the next finer scale are calculated, since they have been calculated in a frame that is moving along with the velocities extracted from coarser scale [1, 5, 28, 18, 8].

Another important aspect of motion estimation is the fact that motion is a dynamic feature of an image sequence. Thus, the longer we observe a movement the more precisely we can estimate and
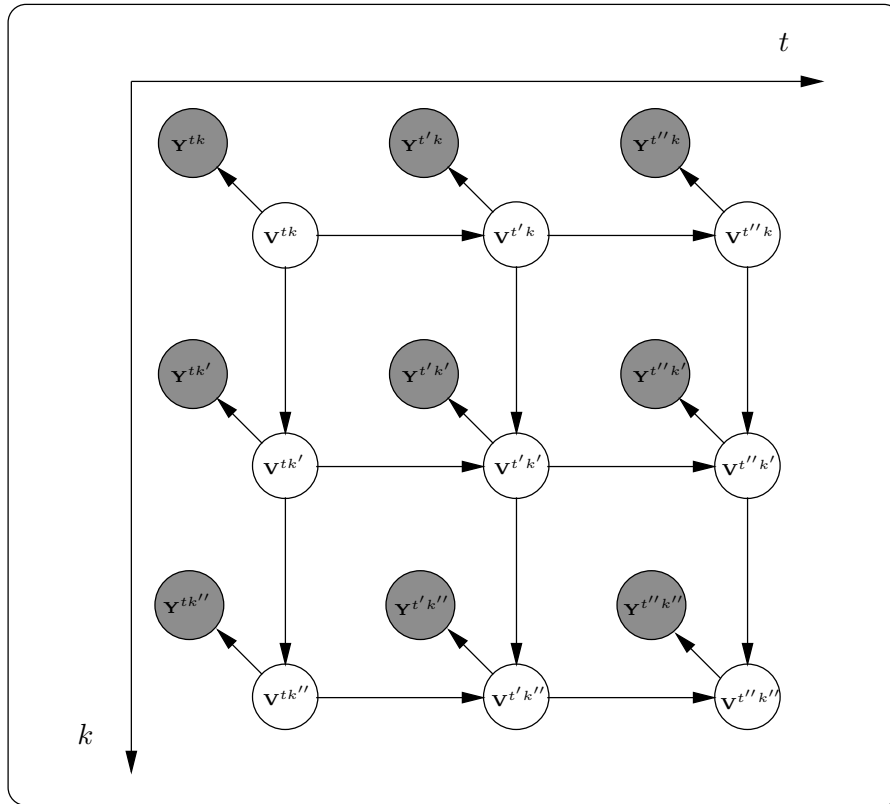
Figure 1.5: A probabilistic directed graphical model for visual motion estimation. Here, $t' = t + 1$ and $t'' = t + 2$ denote future timesteps and $k' = k + 1$ and $k'' = k + 2$ denote finer scales. Observable nodes are shaded gray, hidden nodes are white.

predict its characteristics. This has motivated several approaches [26, 7, 10] to recursively estimate the optical flow over time including a prediction model that defines some temporal relation between pixel movements (see Fig. 1.4 third row). For prediction, a model for the underlying dynamics is needed to predict image motion.

Some authors work in a probabilistic framework assuming that velocity distributions are Gaussian parameterized by a mean and covariance. Kalman filtering can then be used to properly combine the information from scale to scale or time to time taking into account uncertainties of the measurements [26, 21]. The presumption of Gaussian distributed velocity measurements is sometimes incomplete because velocity distributions are often multimodal or ambiguous [22, 30], especially at motion boundaries. To circumvent this problem, particle filtering methods for non-Gaussian velocity distributions have recently been used to improve motion estimation for tracking single or multiple objects in a scene [19, 12].

1.4 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 quite general probabilistic solution which is still computationally tractable.

1.4.1 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 1.5 to model the dynamics of visual motion. The structure of the graphical model in figure 1.5 is similar to a Markov Random field. The difference is that the edges are directed. As can be seen, 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 [34].

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 $t' = t+1$ and $k' = k+1$ 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 $I_{\mathbf{x}}^{tk}$ 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 $\mathbf{v}_{\mathbf{x}}^{tk}$ at each image position $\mathbf{x}$.

1.4.2 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 $k = 1$

$$P(\mathbf{V}^{11}) = \prod_{\mathbf{x}} P(\mathbf{v}_{\mathbf{x}}^{11}), \tag{1.11}$$

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 [30]. 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 $k \in \mathcal{K}$

$$P(\mathbf{Y}^{tk}|\mathbf{V}^{tk}) = \prod_{\mathbf{x}} \ell(\mathbf{Y}^{tk}, \mathbf{v}_{\mathbf{x}}^{tk}). \tag{1.12}$$

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 $\mathbf{V}^{t1}$.

$$
\begin{aligned}
P(\mathbf{V}^{1k'}|\mathbf{V}^{1k}) &= \prod_{\mathbf{x}} \phi_k(\mathbf{v}_{\mathbf{x}}^{1k'}, \mathbf{V}^{1k}), \\
P(\mathbf{V}^{t'1}|\mathbf{V}^{t1}) &= \prod_{\mathbf{x}} \phi_t(\mathbf{v}_{\mathbf{x}}^{t'1}, \mathbf{V}^{t1}), \\
P(\mathbf{V}^{t'k'}|\mathbf{V}^{t'k}, \mathbf{V}^{tk'}) &= \prod_{\mathbf{x}} \phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k})\phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}).
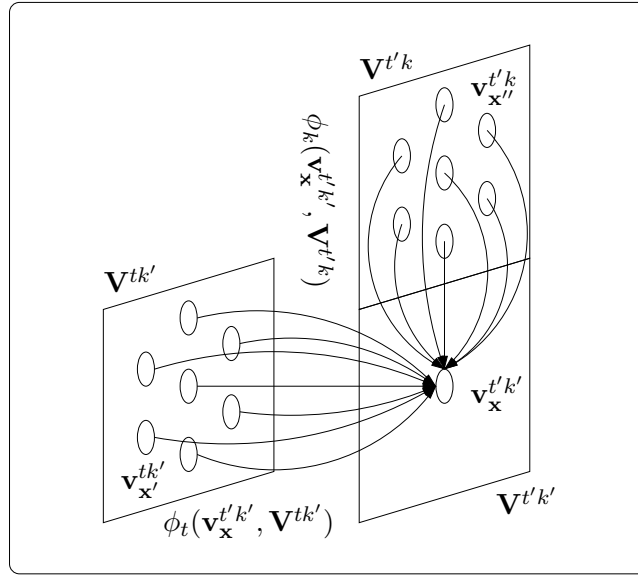\end{aligned}
\tag{1.13}
$$

Figure 1.6: Pairwise potentials of the scale-time transition probability.

These equations explicitly express that the probability distribution for each flow field factorises 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 $P(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{V}^{t'k}, \mathbf{V}^{tk'})$ 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.

1.4.3   A general class of flow field transitions

To further specify the generative model we have to define the formulas for the prior (1.11), the observation likelihood (1.12), and the transitions (1.13). Some concrete examples are given in the next section 1.5. For the flow field transitions in equation (1.13) we propose a certain class of transition probability functions which satisfy a probability mixture model. Equation (1.13) 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 $t \to t'$ 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 $k \to k'$ at every time t'. A sketch of the information flow is shown in figure 1.6.

To motivate the temporal transition factor $\phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'})$ we assume that the direction and speed of a flow vector $\mathbf{v}_{\mathbf{x}}^{t'k'}$ at position $\mathbf{x}$ at time t' is functionally related to a previous flow vector $\mathbf{v}_{\mathbf{x'}}^{tk'}$ at some corresponding position $\mathbf{x}'$ at time t,

$$\mathbf{v}_{\mathbf{x}}^{t'k'} \sim f_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x'}}^{tk'}; \theta_t) \ , \tag{1.14}$$

including some free parameters θ_t that allow for adaptation of the temporal relation. 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

$$\mathbf{x}' \sim f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}; \theta_{xt}) \ . \tag{1.15}$$

In principle f_{xt} can be any arbitrary function that defines the relation between neighboring velocities. Again the free parameters θ_{xt} allow for adaptation of the spatial relation. Note that here we use $\mathbf{v}_{\mathbf{x}}^{t'k'}$ to retrieve the previous corresponding point $\mathbf{x}'$. This is a suitable approximation as long as the similarity $\mathbf{v}_{\mathbf{x}}^{t'k'} \approx \mathbf{v}_{\mathbf{x}'}^{tk'}$ is not heavily violated. Combining both factors (1.14) and (1.15) and integrating $\mathbf{x}'$ leads to the first pairwise potential

$$\phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) = \sum_{\mathbf{x}'} f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}; \theta_{xt}) f_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{tk'}; \theta_t). \tag{1.16}$$

Equivalent to (1.14) for the scale transition factor $\phi_k(\mathbf{v}_{\mathbf{x}}^{1k'}, \mathbf{V}^{1k})$ we assume that the origin of a 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 $\mathbf{x}''$,

$$\mathbf{v}_{\mathbf{x}}^{t'k'} \sim f_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}''}^{t'k}; \theta_k). \tag{1.17}$$

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 (1.15)

$$\mathbf{x}'' \sim f_{xk}(\mathbf{x}'', \mathbf{x}; \theta_{xk}). \tag{1.18}$$

The considerations for the scale transition are analogous to the ones for the temporal transition. Again, combining both factors (1.17) and (1.18) and integrating $\mathbf{x}''$ we get the second pairwise potential

$$\phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) = \sum_{\mathbf{x}''} f_{xk}(\mathbf{x}'', \mathbf{x}; \theta_{xk}) f_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}''}^{t'k}; \theta_k), \tag{1.19}$$

that imposes a spatial smoothness constraint on the flow field via spatial weighting of motion estimations from coarser scale. The combination of both potentials (1.16) and (1.19) results in the complete conditional flow field transition probability as given in (1.13). The transition factors (1.16) and (1.19) allow us to *unroll* two different kinds of spatial constraints along the temporal and the scale axes 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 (1.16) and (1.19) realizes a combination of (A) scale-time prediction and (B) an integration of motion information neighboring in time, in space, and in scale.

1.4.4   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 in figure 1.5

$$\begin{aligned} P(\mathbf{Y}^{1:T,1:K}, \mathbf{V}^{1:T,1:K}) \;=\; & \prod_{t=1}^{T} \prod_{k=1}^{K} P(\mathbf{Y}^{tk}|\mathbf{V}^{tk}) \times P(\mathbf{V}^{11}) \times \\ & \prod_{t=1}^{T-1} P(\mathbf{V}^{t'1}|\mathbf{V}^{t1}) \prod_{k=1}^{K-1} P(\mathbf{V}^{1k'}|\mathbf{V}^{1k}) P(\mathbf{V}^{t'k'}|\mathbf{V}^{t'k}, \mathbf{V}^{tk'}). \end{aligned} \tag{1.20}$$

What we are usually interested in is the probability for some flow field given all the data aquired so far. For the offline case where all the data of a sequence is accessible this would be the probability $P(\mathbf{V}^{tk}|\mathbf{Y}^{1:T,1:K})$. For the online case where only the past data is accessible 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

$$P(\mathbf{V}^{tk}|\mathbf{Y}^{1:T,1:K}) = \sum_{\mathbf{V}^{1:T,1:K}\backslash\mathbf{V}^{tk}} \frac{P(\mathbf{Y}^{1:T,1:K},\mathbf{V}^{1:T,1:K})}{P(\mathbf{Y}^{1:T,1:K})}. \tag{1.21}$$

The online case neglects future observations and simplifies to

$$P(\mathbf{V}^{tk}|\mathbf{Y}^{1:t,1:K}) = \sum_{\mathbf{V}^{1:t,1:K}\backslash\mathbf{V}^{tk}} \frac{P(\mathbf{Y}^{1:t,1:K},\mathbf{V}^{1:t,1:K})}{P(\mathbf{Y}^{1:t,1:K})}. \tag{1.22}$$

Either for the online or offline case, the direct computation of the marginals using equation (1.21) or (1.22) would take exponential time [34]. The most prominent solution to this problem is *Belief Propagation* (BP) which is a very 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 1.5).

1.4.5   Approximate inference

Here, we propose an approximate inference algorithm based on Belief Propagation and restrict ourselves to the online case (1.22) since its extension to the offline case is straightforward [6] (see also section 1.5). 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

$$\alpha(\mathbf{v}_{\mathbf{x}}^{tk}) \approx P(\mathbf{v}_{\mathbf{x}}^{tk}|\mathbf{Y}^{1:t,1:K}). \tag{1.23}$$

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

$$\alpha(\mathbf{v}_{\mathbf{x}}^{11}) = P(\mathbf{v}_{\mathbf{x}}^{11}|\mathbf{Y}^{11}) = \frac{\ell(\mathbf{Y}^{11},\mathbf{v}_{\mathbf{x}}^{11})P(\mathbf{v}_{\mathbf{x}}^{11})}{P(\mathbf{Y}^{11})}. \tag{1.24}$$

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 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.

The factored observation likelihood and the transition probability we introduced in (1.12) and (1.13) ensure that the forward propagated joint belief

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

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.

$$P(\mathbf{V}^{t'k}, \mathbf{V}^{tk'}|\mathbf{Y}^{1:t',1:k'} \setminus \mathbf{Y}^{t'k'}) = P(\mathbf{V}^{t'k}|\mathbf{Y}^{1:t',1:k})P(\mathbf{V}^{tk'}|\mathbf{Y}^{1:t,1:k'}) = \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{t'k})\alpha(\mathbf{v}_{\mathbf{x}}^{tk'}), \qquad (1.25)$$

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 (1.25) via multiplying with the scale-time transition (1.13) and marginalizing over $\mathbf{V}^{t'k}$ and $\mathbf{V}^{tk'}$. The result is multiplied with the new observation likelihood (1.12) and normalized by $P(\mathbf{Y}^{t'k'})$ to get the updated belief

$$
\begin{aligned}
P(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{Y}^{1:t',1:k'}) &= \tfrac{1}{P(\mathbf{Y}^{t'k'})} \quad \ell(\mathbf{Y}^{t'k'}, \mathbf{v}_{\mathbf{x}}^{t'k'}) \sum_{\mathbf{V}^{t'k}} \sum_{\mathbf{V}^{tk'}} P(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{V}^{t'k}, \mathbf{V}^{tk'}) P(\mathbf{V}^{t'k}, \mathbf{V}^{tk'}|\mathbf{Y}^{1:t',1:k'} \setminus \mathbf{Y}^{t'k'}), \\
\alpha(\mathbf{v}_{\mathbf{x}}^{t'k'}) &\propto \quad m_{Y \to v}(\mathbf{v}_{\mathbf{x}}^{t'k'}) \sum_{\mathbf{V}^{t'k}} \sum_{\mathbf{V}^{tk'}} \overbrace{\phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k})\phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'})} \overbrace{\prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{t'k})\alpha(\mathbf{v}_{\mathbf{x}}^{tk'})}, \\
&\propto \quad m_{Y \to v}(\mathbf{v}_{\mathbf{x}}^{t'k'}) \underbrace{\sum_{\mathbf{V}^{t'k}} \phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{t'k})} \underbrace{\sum_{\mathbf{V}^{tk'}} \phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{tk'})}, \\
&\propto \quad m_{Y \to v}(\mathbf{v}_{\mathbf{x}}^{t'k'}) \quad \times \quad m_{k \to k'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) \quad \times \quad m_{t \to t'}(\mathbf{v}_{\mathbf{x}}^{t'k'}). \quad (1.26)
\end{aligned}
$$

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,

$$\alpha(\mathbf{v}_x^{tk}) \propto m_{Y \to v}(\mathbf{v}_{\mathbf{x}}^{tk}) \, m_{k \to k'}(\mathbf{v}_{\mathbf{x}}^{tk}) \, m_{t \to t'}(\mathbf{v}_{\mathbf{x}}^{tk}). \qquad (1.27)$$

Inserting the proposed class of temporal transitions (1.16) into (1.26) leads to the following temporal message

$$
\begin{aligned}
m_{t \to t'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) &= \sum_{\mathbf{V}^{tk'}} \phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{tk'}), \\
&= \sum_{\mathbf{V}^{tk'}} \sum_{\mathbf{x}'} f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}; \theta_{xt}) f_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{tk'}; \theta_t) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{tk'}), \\
&= \sum_{\mathbf{v}_{\mathbf{x}}^{tk'}} \sum_{\mathbf{x}'} f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}; \theta_{xt}) f_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{tk'}; \theta_t) \alpha(\mathbf{v}_{\mathbf{x}}^{tk'}) \underbrace{\sum_{\substack{\mathbf{V}^{tk'} \\ \setminus \mathbf{v}_{\mathbf{x}'}^{tk'}}} \prod_{\mathbf{z} \neq \mathbf{x}'} \alpha(\mathbf{v}_{\mathbf{z}}^{tk'})}_{1}, \\
&= \sum_{\mathbf{x}'} f_{xt}(\mathbf{x}', \mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}; \theta_{xt}) \sum_{\mathbf{v}_{\mathbf{x}}^{tk'}} f_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{tk'}; \theta_t) \alpha(\mathbf{v}_{\mathbf{x}'}^{tk'}). \qquad (1.28)
\end{aligned}
$$

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 $\sum_{\mathbf{V}^{tk'} \backslash \mathbf{v}_{\mathbf{x}'}^{tk'}}$ over all other flow field vectors at $\mathbf{x} \neq \mathbf{x}'$. Then, we use the equivalence $\sum_{\mathbf{V}^{tk'} \backslash \mathbf{v}_{\mathbf{x}'}^{tk'}} \prod_{\mathbf{z} \neq \mathbf{x}'} \alpha(\mathbf{v}_{\mathbf{z}}^{tk'}) = \prod_{\mathbf{z} \neq \mathbf{x}'} \sum_{\mathbf{v}_{\mathbf{z}}^{tk'}} \alpha(\mathbf{v}_{\mathbf{z}}^{tk'}) = 1$. Similarly, we arrive at the scale message if we insert the scale transition (1.19) into (1.26)

$$
\begin{aligned}
m_{k \to k'}(\mathbf{v}_{\mathbf{x}}^{t'k'}) & = \sum_{\mathbf{V}^{t'k}} \phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{t'k}), \\
& = \sum_{\mathbf{V}^{t'k}} \sum_{\mathbf{x}'} f_{xk}(\mathbf{x}', \mathbf{x}; \theta_{xk}) f_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{t'k}; \theta_k) \prod_{\mathbf{x}} \alpha(\mathbf{v}_{\mathbf{x}}^{t'k}), \\
& = \sum_{\mathbf{v}_{\mathbf{x}}^{t'k}} \sum_{\mathbf{x}'} f_{xk}(\mathbf{x}', \mathbf{x}; \theta_{xk}) f_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{t'k}; \theta_k) \alpha(\mathbf{v}_{\mathbf{x}'}^{t'k}) \underbrace{\sum_{\substack{\mathbf{V}^{t'k} \\ \backslash \mathbf{v}_{\mathbf{x}}^{t'k}}} \prod_{\mathbf{z} \neq \mathbf{x}'} \alpha(\mathbf{v}_{\mathbf{z}}^{t'k})}_{1}, \\
& = \sum_{\mathbf{x}'} f_{xk}(\mathbf{x}', \mathbf{x}; \theta_{xk}) \sum_{\mathbf{v}_{\mathbf{x}}^{t'k}} f_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{v}_{\mathbf{x}'}^{t'k}; \theta_k) \alpha(\mathbf{v}_{\mathbf{x}'}^{t'k}). \quad (1.29)
\end{aligned}
$$

Finally, the three equations (1.27), (1.28), and (1.29) define a very efficient tightly coupled scale-time forward filter for visual motion estimation. It realizes a complete probabilistic recurrent estimation of a set of flow fields $\mathbf{V}^{t,1:K}$ with different resolutions k swept along the time dimension t. It follows the principle that the longer you observe a scene and the finer the resolution of the data is the more accurate the flow can be estimated.

1.5   Filter realisations

The pseudo-code, Algorithm 1, shows the very compact form of the derived scale-time filter suitable for an algorithmic implementation. What remains to be done, is the specification of the observation likelihood (1.12) and the potentials of the transition probability (1.16) and (1.19). Without loss of generalization the derivation for the filter assumes discrete state variables which is reflected in using summations $\sum$ for marginalization. If continuous state variables are given the summations $\sum$ simply have to be replaced by integrals $\int$. Everything else keeps being the same. To show the applicability of the framework, we derive two realizations: One for continuous Gaussian and one for discrete grid-based observation likelihoods as well as Mixture of Gaussians and Mixture of Student's-t-distributions transitions. Both realizations have already been published at the International Conference on Machine Learning and Applications [31, 32]. Here, we summarize the essentials of the modelling in relation to the general filter framework. For optical flow estimation results, discussions on the parameters, and benchmark tests we refer to the published material.

1.5.1   The Gaussian 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.

Algorithm 1 Scale-Time Filter

Initialize the priors $\alpha(\mathbf{v}_{\mathbf{x}}^{0,1:K})$

for $t' = 1$ to T **do**
  for $k' = 1$ to K **do**
    for $\mathbf{x} = 1$ to $X^{k'}$ **do**

      Compute the messages

$$m_{Y \to v}(\mathbf{v}_{\mathbf{x}}^{t'k'})$$

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

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

      Update the beliefs

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

    end for
  end for
end for

Gaussian observation likelihood We follow a similar argumentation as Simoncelli *et al.* [22] to obtain the $\ell(\mathbf{v}_{\mathbf{x}}^{tk})$-factors (1.12) 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 [17]. Additionally, we assume i.i.d. additive Gaussian noise s_t, $\mathbf{S}_v$ on the temporal derivatives and the flow field, respectively.

$$\ell(\mathbf{v}_{\mathbf{x}}^{tk}) = \mathcal{N}(-\mathbf{I}_{t,\mathbf{x}}^{tk} | (\nabla \mathbf{I}_{\mathbf{x}}^{tk})^T \mathbf{v}_{\mathbf{x}}^{tk}, \boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk}) , \tag{1.30}$$

$$\boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk} = \begin{pmatrix} \ddots & \ldots & \mathbf{0} \\ \vdots & \sigma_{\ell,\mathbf{xx}'}^{tk} & \vdots \\ \mathbf{0} & \ldots & \ddots \end{pmatrix} , \tag{1.31}$$

$$\sigma_{\ell,\mathbf{xx}'}^{tk} = \frac{(\nabla I_{\mathbf{x}'}^{tk})^T \mathbf{S}_v \nabla I_{\mathbf{x}'}^{tk} + s_t}{f_\ell(\mathbf{x}', \mathbf{x}, t, k)} . \tag{1.32}$$

In notation (1.30), the patches can be regarded as vectors and the covariance matrix $\boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk}$ is a diagonal with entries $\sigma_{\ell,\mathbf{xx}'}^{tk}$ 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 [21] or an anisotropic and inhomogenous Gaussian weighting $f_\ell = \mathcal{N}(\mathbf{x}'|\mathbf{x}, \boldsymbol{\Sigma}_{l,\mathbf{x}}^{tk})$ which is investigated in [32].

In contrast to [21], 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 $\boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk}$ 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 (1.14) we now chose a Gaussian

$$\mathbf{v}_{\mathbf{x}}^{t'k'} \sim \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}'}^{tk'}, \sigma_t) \ , \tag{1.33}$$

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 (1.15) an inhomogeneous anisotropic Gaussian is assumed

$$\mathbf{x}' \sim \mathcal{N}(\mathbf{x}'|\mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}, \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk}) \ . \tag{1.34}$$

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 (1.33) and (1.34) and integrating $\mathbf{x}'$ we get a *Mixture of Gaussians* (MoG) as the first pairwise potential (1.16)

$$\phi_t(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{tk'}) = \sum_{\mathbf{x}'} \mathcal{N}(\mathbf{x}'|\mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t'k'}, \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk}) \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}'}^{tk'}, \sigma_t) \ , \tag{1.35}$$

with the Gaussian spatial coherence constraint being the mixing coefficients. Equivalent to (1.33) for the scale transition factor (1.19) we chose a Gaussian

$$\mathbf{v}_{\mathbf{x}}^{t'k'} \sim \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}''}^{t'k}, \sigma_k) \ , \tag{1.36}$$

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 (1.34)

$$\mathbf{x}'' \sim \mathcal{N}(\mathbf{x}''|\mathbf{x}, \boldsymbol{\Sigma}_{k,\mathbf{x}}^{tk}) \ . \tag{1.37}$$

Again, combining both factors (1.36) and (1.37) and integrating $\mathbf{x}''$ we get a MoG as the second pairwise potential

$$\phi_k(\mathbf{v}_{\mathbf{x}}^{t'k'}, \mathbf{V}^{t'k}) = \sum_{\mathbf{x}''} \mathcal{N}(\mathbf{x}''|\mathbf{x}, \boldsymbol{\Sigma}_{k,\mathbf{x}}^{tk}) \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{t'k'}|\mathbf{v}_{\mathbf{x}''}^{t'k}, \sigma_k) \ , \tag{1.38}$$

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 (1.16) and (1.19) results in the complete conditional flow field transition probability as given in (1.13).

Approximate inference   To arrive at a Gaussian belief we introduce a last approximative restriction. We want every factor of the posterior probability (1.27) to be Gaussian distributed

$$\alpha(\mathbf{v}_x^{tk}) \propto m_{Y \to v}(\mathbf{v}_{\mathbf{x}}^{tk}) \ m_{t \to t'}(\mathbf{v}_{\mathbf{x}}^{tk}) \ m_{k \to k'}(\mathbf{v}_{\mathbf{x}}^{tk}) :\approx \mathcal{N}(\mathbf{v}_{\mathbf{x}}^{tk}|\boldsymbol{\mu}_{\mathbf{x}}^{tk}, \boldsymbol{\Sigma}_{\mathbf{x}}^{tk}) \ . \tag{1.39}$$

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 (1.28, 1.29) leads to two different MoGs for the resulting messages

$$m_{t \to t'}(\mathbf{v}_\mathbf{x}^{t'k'}) = \sum_{\mathbf{x}'} \hat{p}_{\mathbf{x}'}^{t'k'} \mathcal{N}(\mathbf{v}_\mathbf{x}^{t'k'}|\hat{\boldsymbol{\mu}}_{\mathbf{x}'}^{t'k'}, \hat{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{t'k'}) \approx \mathcal{N}(\mathbf{v}_\mathbf{x}^{t'k'}|\boldsymbol{\omega}_\mathbf{x}^{t'k'}, \boldsymbol{\Omega}_\mathbf{x}^{t'k'}) , \tag{1.40}$$

with

$$\hat{p}_{\mathbf{x}'}^{t'k'} = \mathcal{N}(\mathbf{x} - \mathbf{x}'|\boldsymbol{\mu}_{\mathbf{x}'}^{tk'}, \check{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{tk'}) , \tag{1.41}$$

$$\hat{\boldsymbol{\mu}}_{\mathbf{x}'}^{t'k'} = (\sigma_t + \boldsymbol{\Sigma}_{\mathbf{x}'}^{tk'})\check{\Lambda}_{\mathbf{x}'}^{tk'}(\mathbf{x} - \mathbf{x}') + \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk}\check{\Lambda}_{\mathbf{x}'}^{tk'}\boldsymbol{\mu}_{\mathbf{x}'}^{tk'} , \tag{1.42}$$

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}'}^{t'k'} = \boldsymbol{\Sigma}_{t,\mathbf{x}}^{tk}\check{\Lambda}_{\mathbf{x}'}^{tk'}(\sigma_t + \boldsymbol{\Sigma}_{\mathbf{x}'}^{tk'}) , \tag{1.43}$$

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

and

$$m_{k \to k'}(\mathbf{v}_\mathbf{x}^{t'k'}) = \sum_{\mathbf{x}''} \overline{p}_{\mathbf{x}''}^{t'k'} \mathcal{N}(\mathbf{v}_\mathbf{x}^{t'k'}|\boldsymbol{\mu}_{\mathbf{x}''}^{t'k}, \overline{\boldsymbol{\Sigma}}_{\mathbf{x}''}^{t'k'}) \approx \mathcal{N}(\mathbf{v}_\mathbf{x}^{t'k'}|\boldsymbol{\pi}_\mathbf{x}^{t'k'}, \boldsymbol{\Pi}_\mathbf{x}^{t'k'}) , \tag{1.44}$$

with

$$\overline{p}_{\mathbf{x}''}^{t'k'} = \mathcal{N}(\mathbf{x}''|\mathbf{x}, \boldsymbol{\Sigma}_{k,\mathbf{x}}^{tk}) , \quad \overline{\boldsymbol{\Sigma}}_{\mathbf{x}''}^{t'k'} = \sigma_k + \boldsymbol{\Sigma}_{\mathbf{x}''}^{t'k} . \tag{1.45}$$

In order to satisfy the Gaussian constraint formulated in (1.39) the MoG's are collapsed into single Gaussians (1.40, 1.44) 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

$$\alpha(\mathbf{v}_x^{tk}) = \ell(\mathbf{v}_x^{tk}) \, \mathcal{N}(\mathbf{v}_\mathbf{x}^{tk}|\tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk}, \tilde{\boldsymbol{\Sigma}}_\mathbf{x}^{tk}) , \tag{1.46}$$

$$\tilde{\boldsymbol{\Sigma}}_\mathbf{x}^{tk} = \boldsymbol{\Pi}_\mathbf{x}^{tk} \left[\boldsymbol{\Pi}_\mathbf{x}^{tk} + \boldsymbol{\Omega}_\mathbf{x}^{tk}\right]^{-1} \boldsymbol{\Omega}_\mathbf{x}^{tk} , \tag{1.47}$$

$$\tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk} = \boldsymbol{\Omega}_\mathbf{x}^{tk} \left[\boldsymbol{\Pi}_\mathbf{x}^{tk} + \boldsymbol{\Omega}_\mathbf{x}^{tk}\right]^{-1} \boldsymbol{\pi}_\mathbf{x}^{tk} + \boldsymbol{\Pi}_\mathbf{x}^{tk} \left[\boldsymbol{\Pi}_\mathbf{x}^{tk} + \boldsymbol{\Omega}_\mathbf{x}^{tk}\right]^{-1} \boldsymbol{\omega}_\mathbf{x}^{tk} . \tag{1.48}$$

By applying the approximation steps (1.39, 1.40) and (1.44) we guarantee the posterior (1.27) to be Gaussian which allows for Kalman-filter like update equations since the observation is defined to factorize into Gaussian factors (1.30). The final recurrent motion estimation is given by

$$\alpha(\mathbf{v}_\mathbf{x}^{tk}) = \mathcal{N}(\mathbf{v}_\mathbf{x}^{tk}|\boldsymbol{\mu}_\mathbf{x}^{tk}, \boldsymbol{\Sigma}_\mathbf{x}^{tk}) \tag{1.49}$$

$$= \mathcal{N}(-\mathbf{I}_{t,\mathbf{x}}^{tk}|(\nabla \mathbf{I}_\mathbf{x}^{tk})^T \mathbf{v}_\mathbf{x}^{tk}, \boldsymbol{\Sigma}_{\ell,\mathbf{x}}^{tk})\mathcal{N}(\mathbf{v}_\mathbf{x}^{tk}|\tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk}, \tilde{\boldsymbol{\Sigma}}_\mathbf{x}^{tk}) , \tag{1.50}$$

$$\boldsymbol{\Sigma}_\mathbf{x}^{tk} = \left[\tilde{\Lambda}_\mathbf{x}^{tk} + \nabla \mathbf{I}_\mathbf{x}^{tk} \Lambda_{\ell,\mathbf{x}}^{tk} (\nabla \mathbf{I}_\mathbf{x}^{tk})^T\right]^{-1} , \tag{1.51}$$

$$\boldsymbol{\mu}_\mathbf{x}^{tk} = \tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk} - \boldsymbol{\Sigma}_\mathbf{x}^{tk} \nabla \mathbf{I}_\mathbf{x}^{tk} \Lambda_{\ell,\mathbf{x}}^{tk} \tilde{\mathbf{I}}_{t,\mathbf{x}}^{tk} . \tag{1.52}$$

For reasons explained in [21] the innovations process is approximated as the following

$$\tilde{\mathbf{I}}_{t,\mathbf{x}}^{tk} \approx \partial/\partial t \mathcal{T} \left(\mathbf{I}_\mathbf{x}^{tk}, \tilde{\boldsymbol{\mu}}_\mathbf{x}^{tk}\right) , \tag{1.53}$$

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 (1.48). We end up with a Gaussian scale-time filter which is, in comparison to existent filtering approaches [10], [21], [26], not a Kalman Filter realization but 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.

1.5.2 The grid-based realisation

A grid-based filter allows only a discrete set of state variables but is not restricted to any distribution. Here, we neglect the scale dimension but show how past and future observables can be processed offline via a Two-Filter.

Observation likelihood   Now, we define the observation likelihood $P(\mathbf{Y}^t|\mathbf{V}^t)$ by assuming that the likelihood factor $\ell(\mathbf{Y}^t|\mathbf{v}_\mathbf{x}^t)$ of a local velocity $\mathbf{v}_\mathbf{x}^t$ should be related to finding the same or similar image patch centered around $\mathbf{x}$ at time t' that was present at time t but centered around $\mathbf{x} - \mathbf{v}_\mathbf{x}^t$. More rigorously, let $\mathcal{S}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ be the Student's t-distribution and $\mathcal{N}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \lim_{\nu \to \infty} \mathcal{S}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ be the normal distribution of a variable $\mathbf{x}$ with mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$ and the degrees of freedom ν. In the following the covariance is chosen to be isotropic $\boldsymbol{\Sigma} = \sigma^2 \mathbf{E}$ (with identity matrix $\mathbf{E}$). We define

$$\ell(\mathbf{Y}^t|\mathbf{v}_\mathbf{x}^t) = \sum_{\mathbf{x}'} \mathcal{N}(\mathbf{x}'|\mathbf{x}, \varrho_I)\mathcal{S}(I_{\mathbf{x}'}^{t'}|I_{\mathbf{x}'-\mathbf{v}_\mathbf{x}^t}^t, \sigma_I, \nu_I) = \sum_{\mathbf{x}'} \mathcal{N}(\mathbf{x}'|\mathbf{x} - \mathbf{v}_\mathbf{x}^t, \varrho_I)\mathcal{S}(I_{\mathbf{x}'+\mathbf{v}_\mathbf{x}^t}^{t'}|I_{\mathbf{x}'}^t, \sigma_I, \nu_I). \quad (1.54)$$

Here, $\mathcal{N}(\mathbf{x}'|\mathbf{x}, \varrho_I)$ implements a Gaussian weighting of locality centered around $\mathbf{x}$ for I^{t+1} and around $\mathbf{x} - \mathbf{v}_\mathbf{x}^t$ 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 $\mathcal{S}(I_{\mathbf{x}'}^{t+1}, I_{\mathbf{x}'-\mathbf{v}_\mathbf{x}^t}^t, \sigma_I, \nu_I)$ 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.

Mixture of Student's t transition   Similarly to equation (1.54), we define the transition probability $P(\mathbf{v}_\mathbf{x}^{t+1}\|\mathbf{V}^t)$ by assuming that the flow field transforms according to itself like defined in 1.16 and further specified as

$$P(\mathbf{v}_\mathbf{x}^{t+1}|\mathbf{V}^t) \propto \mathcal{N}(\mathbf{x}'|\mathbf{x} - \mathbf{v}_\mathbf{x}^{t+1}, \varrho_V)\mathcal{S}(\mathbf{v}_\mathbf{x}^{t+1}|\mathbf{v}_{\mathbf{x}'}^t, \sigma_V, \nu_V). \quad (1.55)$$

Using a heavy tailed Student's t distribution, we assume robust spatiotemporal coherence because evaluations on first derivative optical flow statistics [20] and on prior distributions that allow to imitate human speed discrimination tasks [27] provide strong indication that they resemble such heavy tailed Student's t-distributions. 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 $t + 1$. We introduced new parameters ϱ_V and σ_V for the uncertainty in spatial identification between two images and the transition noise between $\mathbf{V}^t$ and $\mathbf{v}_\mathbf{x}^{t+1}$, respectively. 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 (1.55) equals the transition probability formulated in [9] which expresses the belief that pixels *are, on average, moving along a straight line with constant velocity*. Therefore, the proposed spatiotemporal transition model 1.55 can be seen as a generalization of the transition model proposed by Burgi et al. [9].

Two-Filter inference Like beforehand, for inference we need to propagate beliefs over the flow field $\mathbf{V}^t$. Storing a distribution over a whole flow field $\mathbf{V}^t$ 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 [16] where the "backward filter" is strictly symmetric to the forward filter.

Following the derivation for the temporal belief propagation 1.28 and specifying the transition probability as in equation 1.55 the forward filter reads

$$\alpha(\mathbf{v}_{\mathbf{x}}^{t+1}) \quad \propto \quad \ell(\mathbf{Y}^{t+1}|\mathbf{v}_{\mathbf{x}}^{t+1}) \, \alpha^*(\mathbf{v}_{\mathbf{x}}^{t+1}) \tag{1.56}$$

$$\alpha^*(\mathbf{v}_{\mathbf{x}}^{t+1}) \quad \propto \quad \sum_{\mathbf{x}'} \mathcal{N}(\mathbf{x}'|\mathbf{x} - \mathbf{v}_{\mathbf{x}}^{t+1}, \varrho_V) \sum_{\mathbf{v}_{\mathbf{x}'}^t} \mathcal{S}(\mathbf{v}_{\mathbf{x}}^{t+1}|\mathbf{v}_{\mathbf{x}'}^t, \sigma_V, \nu_V) \, \alpha(\mathbf{v}_{\mathbf{x}'}^t) \; . \tag{1.57}$$

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

$$P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{V}^{t+1}) \propto \sum_{\mathbf{x}'} \mathcal{N}(\mathbf{x}'|\mathbf{x} + \mathbf{v}_{\mathbf{x}}^t, \varrho_V) \, \mathcal{S}(\mathbf{v}_{\mathbf{x}}^t, \mathbf{v}_{\mathbf{x}'}^{t+1}|\sigma_V, \nu_V) \tag{1.58}$$

instead of (1.55). This equation is motivated in exactly the same way as we motivated (1.55): we assume that $\mathbf{v}_{\mathbf{x}}^t \sim \mathcal{S}(\mathbf{v}_{\mathbf{x}'}^{t+1}, \sigma_V, \nu_V)$ for a corresponding position $\mathbf{x}'$ in the subsequent image, and that $\mathbf{x}' \sim \mathcal{N}(\mathbf{x} - \mathbf{v}_{\mathbf{x}}^t, \varrho_V)$ is itself defined by $\mathbf{v}_{\mathbf{x}}^t$. However, note that using this symmetry of argumentation is actually an approximation to our model because applying Bayes rule on (1.55) would lead to a different, non-factored $P(\mathbf{V}^t|\mathbf{V}^{t+1})$. What we gain by the approximation $P(\mathbf{V}^t|\mathbf{V}^{t+1}) \approx \prod_x P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{V}^{t+1})$ are factored β's which are feasible to maintain computationally. The backward filter equations read

$$\beta^*(\mathbf{v}_{\mathbf{x}}^t) \quad \propto \quad \ell(\mathbf{Y}^t|\mathbf{v}_{\mathbf{x}}^t) \, \beta(\mathbf{v}_{\mathbf{x}}^t) \; , \tag{1.59}$$

$$\beta(\mathbf{v}_{\mathbf{x}}^t) \quad \propto \quad \sum_{\mathbf{x}'} \mathcal{N}(\mathbf{x}'|\mathbf{x} + \mathbf{v}_{\mathbf{x}}^t, \varrho_v) \sum_{\mathbf{v}_{\mathbf{x}'}^{t+1}} \mathcal{S}(\mathbf{v}_{\mathbf{x}}^t|\mathbf{v}_{\mathbf{x}'}^{t+1}, \sigma_V, \nu_V) \beta^*(\mathbf{v}_{\mathbf{x}'}^{t+1}) \; . \tag{1.60}$$

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

$$\gamma(\mathbf{v}_{\mathbf{x}}^t) \quad = \quad P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{Y}^{1:T}) = \frac{P(\mathbf{Y}^{t+1:T}|\mathbf{v}_{\mathbf{x}}^t) \, P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{Y}^{1:t})}{P(\mathbf{Y}^{1:T})} \tag{1.61}$$

$$= \quad \frac{P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{Y}^{t+1:T})P(\mathbf{Y}^{t+1:T})P(\mathbf{v}_{\mathbf{x}}^t|\mathbf{Y}^{1:t})}{P(\mathbf{v}_{\mathbf{x}}^t)P(\mathbf{Y}^{1:T})}$$

$$\propto \quad \alpha(\mathbf{v}_{\mathbf{x}}^t) \, \beta(\mathbf{v}_{\mathbf{x}}^t) \, \frac{1}{P(\mathbf{v}_{\mathbf{x}}^t)} \; , \tag{1.62}$$

with $P(\mathbf{Y}^{t+1:T})$ and $P(\mathbf{Y}^{1:T})$ being constant. If both the forward and backward filters are initialized with $\alpha(\mathbf{v}_{\mathbf{x}}^0) = \beta(\mathbf{v}_{\mathbf{x}}^T) = P(\mathbf{v}_{\mathbf{x}})$ we can identify the unconditioned distribution $P(\mathbf{v}_{\mathbf{x}}^t)$ with the prior $P(\mathbf{v}_{\mathbf{x}})$. For details on the standard forward-backward-algorithm we refer to [6].

1.6 Summary

A reliable and robust motion estimate is an important low-level processing unit that has the potential to bootstrap a lot of visual perception tasks to be solved by a cognitive vision system. Since the

estimation of motion information has to rely on highly uncertain visual information a probabilistic treatment of the problem is proposed. Based on three basic approaches to solve motion ambiguities, the derivation of a probabilistic filter is given that combines all these three approaches into one recurrent framework. The derivation comprises an efficient approximate inference algorithm based on belief propagation applied on a directed graphical model with a graph topology suitable for intertwining belief propagation along two dimensions, scale and time, simultaneously. Introducing some factorisation assumptions and a special class of transition probabilities results in a very compact and computationally efficient algorithm. For this algorithm two implementations are presented. The first one realizes a purely factored Gaussian belief propagation and the second one the propagation of a factored non-parametric discrete distribution. The presented framework provides a flexible basis for the realization of user specific motion estimation algorithms with the focus on online applications. It also serves as an exploration platform to investigate in adaptation mechanisms and online learning strategies for example to improve the optical flow estimation accuracy or increase the robustness for highly dynamic scenes.

Bibliography

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *IJCV: International Journal of Computer Vision*, 2(3):283–310, January 1989.

[2] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV)*, 2007.

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

[4] S.S. Beauchemin and J.L. Barron. The computation of optical flow. *ACM Computing Surveys*, 27(3):433–467, September 1995.

[5] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings of the Second European Conference on Computer Vision (ECCV)*, volume 588 of *Lecture Notes In Computer Science*, pages 237–252, London, UK, 1992. Springer-Verlag.

[6] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, 2006.

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

[8] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision (ECCV)*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36, Prague, Czech Republic, May 2004. Springer-Verlag.

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

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

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

[12] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *IJCV: International Journal of Computer Vision*, 29:5–28, 1998.

[13] B. Jähne. *Digitale Bildverarbeitung.* Springer-Verlag Berlin Heidelberg, 1997.

[14] J.J.Gibson. The perception of the visual world. Technical report, Houghton Mifflin Company, Boston, MA, 1950.

[15] S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, Englewood Cliffs, NJ, 1993.

[16] 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.

[17] B. D. Lukas and T. Kanade. An iterative image-registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 674–679, Vancouver, Canada, 1981.

[18] E. Memin and P. Perez. A multigrid approach for hierarchical motion estimation. In *Proceedings of the Sixth International Conference on Computer Vision (ICCV)*, pages 933–938, Bombay, India, January 1998.

[19] Y. Rosenberg and M. Werman. A general filter for measurements with any probability distribution. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 106–111, San Juan, Puerto Rico, June 1997. IEEE Computer Society.

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

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

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

[23] E.P. Simoncelli. *Distributed Representation and Analysis of Visual Motion.* PhD thesis, MIT Department of Electrical Engineering and Computer Science, 1993.

[24] E.P. Simoncelli. Local analysis of visual motion. In *The Visual Neuroscience*. MIT Press, 2003.

[25] A. Singh. An estimation-theoretic framework for image-flow computation. In *Proceedings of the Third International Conference on Computer Vision*, pages 168–177, Osaka, Japan, December 1990.

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

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

[28] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. *IJCV: International Journal of Computer Vision*, 14(1):5–19, 1995.

[29] Y. Weiss. *Bayesian Motion Estimation and Segmentation.* PhD thesis, MIT Department of Brain and Cognitive Science, 1993.

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

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

[32] V. Willert, M. Toussaint, J. Eggert, and E. Körner. Probabilistic exploitation of the lucas and kanade smoothness constraint. In *ICMLA*, 2008.

[33] Qing X. Wu. A correlation-relaxation-labeling framework for computing optical flow - template matching from a new perspective. *IEEE Trans. PAMI*, 17(3):843–853, 1995.

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

[35] J.S. Zelek. Bayesian real-time optical flow. In *Proceedings of the 15th International Conference on Vision Interface (VI)*, pages 310–315, Calgary, Canada, May 2002.

[36] C. Zetzsche and G. Krieger. Nonlinear mechanisms and higher-order statistics in biological vision and electronic image processing: Review and perspectives. *Journal of Electronic Imaging*, 10(1):56–99, 2001.