Non-Gaussian Velocity Distributions Integrated over Space, Time and Scales

Volker Willert, Julian Eggert, Jürgen Adamy, Edgar Körner

2006

Preprint:

This is an accepted article published in IEEE Transactions on Systems, Man and Cybernetics B. The final authenticated version is available online at: https://doi.org/[DOI not available]

Non-Gaussian Velocity Distributions Integrated over Space, Time and Scales

Volker Willert, Julian Eggert, Jürgen Adamy, Edgar Körner

***Abstract*— Velocity distributions are an enhanced representation of image velocity containing more velocity information than velocity vectors. In particular, non-Gaussian velocity distributions allow for the representation of ambiguous motion information caused by the aperture problem or multiple motions at motion boundaries. To resolve motion ambiguities, discrete non-Gaussian velocity distributions are suggested that are integrated over space, time and scales using a joint Bayesian prediction and refinement approach. This leads to a hierarchical velocity distribution representation from which robust velocity estimates for both slow and high speeds as well as statistical confidence measures rating the velocity estimates can be computed.**

***Index Terms*— Velocity Likelihood, Bayesian Tracking, Multiscale Representation, Optical Flow.**

I. Introduction

TRADITIONALLY, motion estimates of moving objects in an image sequence are represented using vector fields consisting of velocity vectors each describing the motion at a particular image region or pixel [1],[2]. Yet in most cases single velocity vector estimates at each image location are an incomplete representation to characterize motion unambiguously which may introduce great errors in subsequent and sequential motion estimations. One reason for the incomplete representation is that the underlying generative model for the motion measurement process is only an approximation for the real 2D image movement/formation. A further reason is that the image data is disturbed by sensor noise.

Besides incomplete models and noisy measurements there are a series of fundamental problems concerning motion estimation. These are the *aperture problem*, the *correspondence problem* within image regions with low contrast (also named the blank wall problem) or periodic texture and the appearance of *multiple motions* caused by occlusions at motion boundaries and transparency of moving objects.

To deal with these ambiguities a higher-dimensional, enhanced velocity representation is proposed. For this purpose, the velocity of an image patch and the image itself are understood as statistical signals. This implies probabilities for the existence of image features, like gray value and velocity, and leads to a conditional probability density function (pdf) in feature and velocity space that can be interpreted as a likelihood function [3],[4]. The expectation is that pdfs are able to tackle the addressed problems related to motion processing, like ambiguous motion, occlusion and transparency. As can be seen in [5],[6],[7] specific information about the mentioned problems can, in principle, be extracted from the shape of the pdfs.

Manuscript received September 30, 2004; revised February 7, 2005.

During the last ten years velocity distributions have been suggested and discussed by several authors mainly using two approaches: the *gradient-based* brightness change constraint equation and the *correlation-based* patch matching technique [4],[8],[9],[10].

One established method to reduce ambiguities is the integration of motion information over space. That means, interactions between neighboring velocities or even higher order derivations are considered [1],[2],[11],[12]. This is often accounted for by smoothness constraints for neighboring velocities assuming that all pixels within the neighborhood move similarly.

Another challenge in motion estimation is to improve and stabilize the estimation over time. This can be done by temporal smoothness constraints and/or temporal prediction algorithms. For prediction, a model for the underlying dynamics is needed to predict image motion.

Further improvements are made using multiscale approaches. This is desirable, e.g., for being able to represent both high and low velocities at good resolutions with a reasonable effort. This is usually done in such a way that the higher velocities at coarser scale are calculated first, then a shifted/warped version of the image is calculated using these higher *absolute* velocities, and afterwards the lower velocities at the next finer scale are calculated. These then correspond to *relative* lower velocities, since they have been calculated in a frame that is moving along with the higher velocities from coarser scale [12],[13],[14], [15],[16].

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 [17],[18]. Like mentioned before, the presumption of Gaussian distributed velocity measurements is sometimes incomplete because velocity distributions are often multimodal or ambiguous [4],[10], 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],[20],[21].

In this work, we propose discrete multimodal pdfs for velocity estimation to allow for a better velocity representation in particular at problematic regions containing occlusion effects and ambiguous information. The main focus of the presented model is to stick on the concept of representing velocity information using velocity distributions (instead of velocity vectors) for all pixels in the image (instead of single object positions) within the entire framework. We introduce a joint space-time integration for non-Gaussian velocity distributions. This approach is extended to reach a joint space-time-scale integration algorithm to reduce motion ambiguities and to estimate image motion also for large displacements over long-time sequences.

First in Sec. II, we give a short biological interpretation of our system structure and the main principles.

In Sec. III a *linear generative model* of image patch formation over time is introduced. Here we assume that the changes in two consecutive images depend on the displacements as well as brightness and contrast variations of localized image patches. The results are contrast and brightness invariant velocity distributions, based on a straight correlation measure comparing windowed image patches of consecutive images. To account for brightness and contrast changes has previously also been proposed in [22], [23] using a gradient-based approach.

In Sec. IV we present an approach for propagation of velocity distributions on the basis of the generative model explained in Sec. III. This is done in a Bayesian manner inspired by *grid based methods* that are able to approximate an optimal Bayesian solution [21], [24].

A model for overall image warping is proposed in Sec. V that includes velocity pdfs. To this end, the local generative model for independent patch movements is extended to a global generative model including patch dependencies which are conditional on the overlap.

Afterwards in Sec. VI, we set up a hierarchical chain of velocity distributions from coarse to fine spatial scale and from larger to smaller relative velocities. At each stage of the pyramid, the distributions for the absolute velocities are improved using the distributions from the coarser spatial scale and the previous timestep of the image sequence. This is done exclusively on the basis of velocity distributions, and is different from other frameworks that operate through several hierarchies but rely on velocity fields when combining information from several hierarchy levels [13], [14].

Finally in Sec. VII some results and comparisons to other approaches are presented.

II. Biologically Inspired System Structure

The system structure for combining velocity information among scales and time is illustrated in Fig. 1. Only two levels of the complete hierarchy are shown for the sake of simplicity. The image pyramid consists of several resolution levels of the input image achieved by the Gaussian decomposition [25]. Every image resolution has its corresponding velocity distribution map. Every column of these maps is a 1D representation of a 2D velocity pdf $\rho(\mathbf{v}|\mathbf{x})$ for the corresponding image location $\mathbf{x}$ with a particular resolution in velocity space $\mathbf{v}$. One column is build up on several spheres with every sphere representing a probability value. This leads to a pyramid of *velocity distribution maps* for every resolution level.

The principle of refinement from coarser to finer scales as well as the resolving of motion ambiguities over time, that are explained in detail in Sec. IV and Sec. VI, can also be seen from a more biological viewpoint. Every hierarchy of the distribution pyramid can be interpreted as a specific brain area consisting of a retinotopically ordered bank of receptive fields tuned to various velocities [24]. The size of the spheres represents the size of the receptive fields. Every velocity distribution map is built from velocity tuned cells where every layer of the map represents a particular velocity $\mathbf{v}$ (direction and speed) for all locations $\mathbf{x}$ of the image (retina). This implies a columnar structure within every hierarchy whereas the columns at each spatial point $\mathbf{x}$ are composed of a set of cells tuned to a range of velocities $\mathbf{v}$. Velocity cells at finer scale with smaller receptive fields are pretuned by velocity cells from coarser scale with larger receptive fields and the cells' previous activities. If 1.) the cell activity of the correspondent spatial locations are consistent and 2.) do agree with the neighboring activities over time, the prediction is reinforced and the uncertainty about the refinement is decreased. To the contrary, inconsistent measurements may increase the uncertainty. In our distribution pyramid, the possibility of refinement and resolving at each level is conditioned by the activities at the previous level and previous timestep. A higher activity at previous level and timestep implies more unimodal distributions and therefore better chances for refinement and resolving.

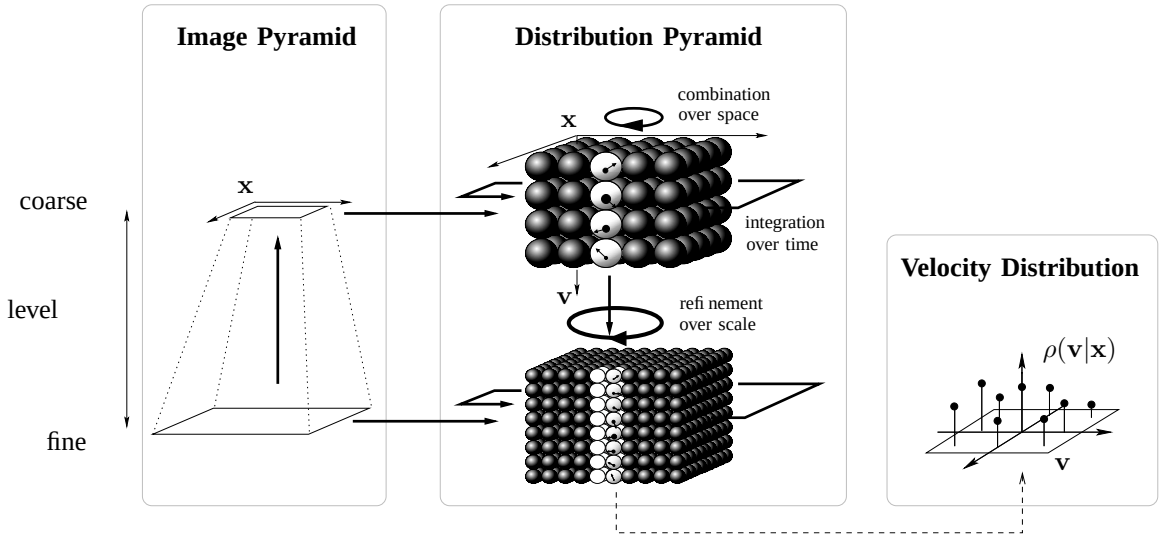

Fig. 1. Biologically inspired hierarchical system structure. The model consists of several processing levels at different spatial resolutions gained from a Gaussian pyramidal decomposition of the input image (left). Every image resolution has its corresponding velocity distribution map (middle). Every column of these 3D maps is a 1D representation of a 2D velocity pdf $\rho(\mathbf{v}|\mathbf{x})$ (shown in the lower right corner) for the corresponding image location $\mathbf{x}$ with a particular resolution in velocity space $\mathbf{v}$. One column is build up on several spheres with every sphere representing a probability value. This leads to a distribution pyramid of *velocity distribution maps* for every resolution level.

The presented architecture combines the advantages of a hierarchical structure, space-time integration and the representation of velocities using distributions. It allows for a coarse-to-fine strategy hand in hand with time propagation of velocity distributions.

III. Velocity Likelihood Formulation

First of all, we introduce the notation used to formulate our model mathematically.

a, A	scalar
$\mathbf{a}$	vector
$\mathbf{A}$	matrix or 2D image
$\mathcal{A}$	function of vectors, matrices or images
$\mathbf{1}$	vector of ones
$\mathbf{1}$	matrix of ones
$\mathbf{A}\odot\mathbf{B}$	componentwise multiplication of two matrices
$\mathbf{A}^{@}$	componentwise exponentiation with α of a matrix

In an image sequence, every image $\mathbf{I}^t$ at time t consists of pixels at locations $\mathbf{x}$. Each pixel is associated with properties like its gray value $G^t_{\mathbf{x}}$ and its velocity vector $\mathbf{v}^t_{\mathbf{x}}$. $\mathbf{G}^t$ denotes the matrix of all gray values of image $\mathbf{I}^t$. The *optical flow* is an approximation of the *motion field* which is the 2D projection of all physical velocities in the 3D world on the corresponding pixel locations $\mathbf{x}$ in the image at a time t. It is usually gained by comparing localized patches of two consecutive images $\mathbf{I}^t$ and $\mathbf{I}^{t+\Delta t}$ with each other. To do this, we define $\mathbf{W}\odot\mathbf{G}^{t,\mathbf{x}}$ as the patch of gray values taken from an image $\mathbf{I}^t$, whereas $\mathbf{G}^{t,\mathbf{x}} := \mathcal{T}^{\{\mathbf{x}\}}\mathbf{G}^t$ are all gray values of image $\mathbf{I}^t$ shifted to $\mathbf{x}$. The shift-operator is defined as follows: $\mathcal{T}^{\{\Delta\mathbf{x}\}}G^t_{\mathbf{x}} := G^t_{\mathbf{x}-\Delta\mathbf{x}}$. The $\mathbf{W}$ defines a window (e.g. a 2-dimensional Gaussian window) that restricts the size of the patch. One possibility to calculate an estimate for the image velocities is to assume that all gray values inside of a patch around $\mathbf{x}$ move with a common velocity $\mathbf{v}^t_{\mathbf{x}}$ for some time Δt, resulting in a displacement of the patch. This basically amounts to a search for correspondences of weighted patches of gray values (displaced with respect to each other) $\mathbf{W}\odot\mathbf{G}^{t+\Delta t,\mathbf{x}+\Delta\mathbf{x}}$ and $\mathbf{W}\odot\mathbf{G}^{t,\mathbf{x}}$ taken from the two images $\mathbf{I}^{t+\Delta t}$ and $\mathbf{I}^t$.

To formulate the calculation of the local motion estimate more precisely, we recur to a linear generative model. Our approximative approach is that for the motion measurement every image patch $\mathbf{W}\odot\mathbf{G}^{t,\mathbf{x}}$ can be considered independently. The assumption of independent motion measurements does not take into consideration statistical correlations which occur for overlapping patches and similar velocities but is useful to simplify the computation of the velocity likelihood [1]. Every image patch $\mathbf{W}\odot\mathbf{G}^{t+\Delta t,\mathbf{x}+\Delta\mathbf{x}}$ is causally linked with its preceding image patch $\mathbf{W}\odot\mathbf{G}^{t,\mathbf{x}}$ in the following way: We assume that an image $\mathbf{I}^t$ patch $\mathbf{W}\odot\mathbf{G}^{t,\mathbf{x}}$ with an associated velocity $\mathbf{v}^t_{\mathbf{x}} = \Delta\mathbf{x}/\Delta t$ is displaced by $\Delta\mathbf{x}$ during time Δt to reappear in image $\mathbf{I}^{t+\Delta t}$ at location $\mathbf{x}+\Delta\mathbf{x}$, so

[1] But we consider spatial correlations in Sec. IV, Eq. (14).

that for this particular patch it is

$$\mathbf{W} \odot \mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}} = \mathbf{W} \odot \mathbf{G}^{t, \mathbf{x}} \quad . \tag{1}$$

In addition, we assume that during this process the gray levels are jittered by noise η, and that brightness and contrast variations may occur over time. The brightness and contrast changes are accounted for by a scaling parameter λ and a bias κ so that we arrive at

$$\mathbf{W} \odot \mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}} = \lambda \mathbf{W} \odot \mathbf{G}^{t, \mathbf{x}} + \kappa \, \mathbf{W} + \eta \, \mathbf{1} \quad . \tag{2}$$

Assuming that the image noise is zero mean Gaussian with variance σ_η^2, the likelihood that $\mathbf{G}^{t,\mathbf{x}}$ is a match for $\mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}}$, given a velocity $\mathbf{v}_{\mathbf{x}}^t$, the window function $\mathbf{W}$ and the parameters λ, κ and σ_η, can be written down as:

$$\rho_{\lambda,\kappa,\sigma_\eta}(\mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}}, \mathbf{G}^{t,\mathbf{x}} | \mathbf{v}_{\mathbf{x}}^t, \mathbf{W}) = \\ = \frac{1}{\sigma_\eta \sqrt{2\pi}} e^{-\frac{1}{2\sigma_\eta^2} \left\| \mathbf{W} \odot \left(\lambda \, \mathbf{G}^{t,\mathbf{x}} + \kappa \, \mathbf{1} - \mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}} \right) \right\|^2} \quad . \tag{3}$$

We now proceed to make Eq. (3) less dependent on λ and κ. For this purpose, we maximize the likelihood Eq. (3) with respect to the scaling and shift parameters, λ and κ. This amounts to minimizing the exponent, so that we want to find

$$\{\lambda^*, \kappa^*\} := \min_{\lambda,\kappa} \left\| \mathbf{W} \odot \left(\lambda \mathbf{G}^{t,\mathbf{x}} + \kappa \mathbf{1} - \mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}} \right) \right\|^2 . \tag{4}$$

This leads to:

$$\lambda^* = \frac{\varrho_{\mathbf{G}^{t,\mathbf{x}}, \mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}}} \cdot \sigma_{\mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}}}}{\sigma_{\mathbf{G}^{t,\mathbf{x}}}} \quad \text{and} \tag{5}$$

$$\kappa^* = \mu_{\mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}}} - \lambda^* \cdot \mu_{\mathbf{G}^{t,\mathbf{x}}} \quad , \quad \text{with} \tag{6}$$

$$\mu_{\mathbf{X}} = \langle \mathbf{X} \rangle := \frac{{}_1{}^T \mathbf{X} \odot \mathbf{W}^{②}{}_1}{{}_1{}^T \mathbf{W}^{②}{}_1} \quad , \tag{7}$$

$$\sigma_{\mathbf{X}}^2 = \langle \mathbf{X}^{②} \rangle - \langle \mathbf{X} \rangle^2 \quad , \quad \text{and} \tag{8}$$

$$\varrho_{\mathbf{X},\mathbf{Y}} = \frac{1}{\sigma_{\mathbf{X}} \cdot \sigma_{\mathbf{Y}}} \left\langle (\mathbf{X} - \mu_{\mathbf{X}} \mathbf{1}) \odot (\mathbf{Y} - \mu_{\mathbf{Y}} \mathbf{1}) \right\rangle \quad , \tag{9}$$

whereas $\mathbf{X}$ and $\mathbf{Y}$ have to be replaced by $\mathbf{G}^{t,\mathbf{x}}$ and $\mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}}$, respectively.

Inserting Eq. (5) and (6) into (3), so that $\lambda = \lambda^*$ and $\kappa = \kappa^*$, leads to the final likelihood formulation:

$$\rho_{\lambda^*,\kappa^*,\sigma_\eta}(\mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}}, \mathbf{G}^{t,\mathbf{x}} | \mathbf{v}_{\mathbf{x}}^t, \mathbf{W}) := \tag{10}$$

$$\rho^t(\mathbf{x}|\mathbf{v}) = \frac{1}{\sigma_\eta \sqrt{2\pi}} e^{-\frac{1}{2} \cdot \left(\frac{\sigma_{\mathbf{G}^{t,\mathbf{x}}}}{\sigma_\eta} \right)^2 \left(1 - \varrho^2_{\mathbf{G}^{t,\mathbf{x}}, \mathbf{G}^{t+\Delta t, \mathbf{x}+\mathbf{v} \cdot \Delta t}} \right)} .$$

To make the notation shorter, we replace $\mathbf{G}^{t+\Delta t, \mathbf{x}+\Delta \mathbf{x}}, \mathbf{G}^{t,\mathbf{x}}$ by $\mathbf{x}$ and drop σ_η and $\mathbf{W}$. The likelihood $\rho^t(\mathbf{x}|\mathbf{v})$ describes the probabilities of the image data at location $\mathbf{x}$ at time t given discrete motion hypotheses $\mathbf{v} = \Delta \mathbf{x} / \Delta t$.

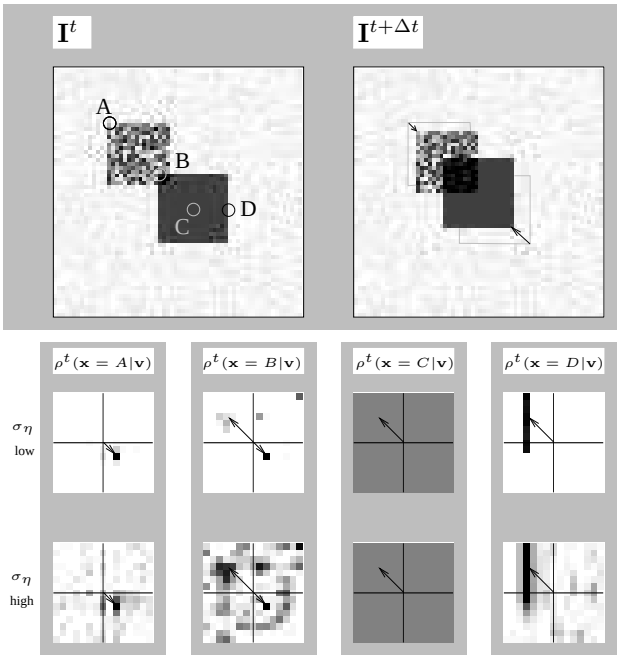

Fig. 2. Influence of the noise parameter σ_η on the likelihood $\rho^t(\mathbf{x}|\mathbf{v})$ calculated using two consecutive frames $\mathbf{I}^t$ and $\mathbf{I}^{t+\Delta t}$. Upper frame: Two squares are moving towards each other with different velocities, whereas the structured square in the upper left is transparent and overlaps the non-structured square in the lower right. Upper frames: Velocity likelihoods (the darker values denote higher probabilities) for a low and high noise parameter σ_η. The likelihood $\rho^t(\mathbf{x} = A|\mathbf{v})$ at location A shows an unimodal distribution for the corner of a square. At location B two motions from both squares are present which is reflected in a multimodal distribution of $\rho^t(\mathbf{x} = B|\mathbf{v})$. At location C there is no structure at all which leads to a completely equally distributed likelihood $\rho^t(\mathbf{x} = C|\mathbf{v})$. The well known aperture problem that occurs at moving edges, like at location D, is represented in an ambiguous distribution along the edge $\rho^t(\mathbf{x} = D|\mathbf{v})$.

Eq. (10) exhibits some additional properties compared to other proposed velocity likelihood measures [4],[9],[10]. Our velocity likelihood is derived from a generative model Eq. (2) that is based on a patch match and allows for local changes in contrast and brightness. This is not accounted for by comparable likelihood formulations that are based on the gradient constraint equation [4],[7]. The noise is assumed to be in the imaging domain with zero mean Gaussian noise added to the image gray values. Other approaches derive noise models in the derivative domain [4],[26].

Our likelihood measure results in a straight correlation method [27] with the weighted empirical correlation coefficient Eq. (9) that ensures that local changes in illumination have minimal influence on the accuracy of the likelihood. Another property of Eq. (10) is given

by the ratio of the variance of the patch at location $\mathbf{x}$ to the variance of the Gaussian distributed noise $\sigma^2_{\mathbf{G}^{t,\mathbf{x}}}/\sigma^2_\eta$. If σ_η is chosen low, then only high values of $\varrho^2_{\mathbf{G}^{t,\mathbf{x}},\mathbf{G}^{t+\Delta t,\mathbf{x}+\Delta\mathbf{x}}}$ contribute to the distribution and less contrastive patches tend to have a higher impact on the resulting likelihood distribution. For higher noise level σ_η, more contrastive patches are needed to get a significantly peaked distribution. The influence of the noise parameter σ_η on the likelihood $\rho^t(\mathbf{x}|\mathbf{v})$ and some typical ambiguous motion situations are shown in Fig. 2.

IV. Integration over Space-Time

With the aim to develop a joint integration of velocity distributions over space-time and scales we first introduce the propagation over space and time algorithm which is then combined with the integration over scales strategy in section VI.

To get velocity pdfs out of the observed likelihood measures $\rho^t(\mathbf{x}|\mathbf{v})$ Bayesian Inference is used to find the probabilities of hypotheses $\rho^t(\mathbf{v}|\mathbf{x})$ given observations. Here, $\mathbf{x}$ denotes the observation that the patch $\mathbf{W}\odot\mathbf{G}^{t,\mathbf{x}}$ matches the patch $\mathbf{W}\odot\mathbf{G}^{t+\Delta t,\mathbf{x}+\Delta\mathbf{x}}$ which has been explained in Sec. III. This is a common way to estimate the posterior pdfs $\rho^t(\mathbf{v}|\mathbf{x})$ that hold the probabilities of the motion hypotheses $\mathbf{v}$ given the image data [10], [19]. Applying Bayes' rule a prior $\rho(\mathbf{v})$ is combined with the likelihood to calculate the posterior

$$\rho^t(\mathbf{v}|\mathbf{x}) \sim \rho^t(\mathbf{x}|\mathbf{v})\rho(\mathbf{v}) \quad . \tag{11}$$

The prior $\rho(\mathbf{v})$ is a common velocity distribution for all positions $\mathbf{x}$ that assumes probabilities for motion hypotheses v in the absence of any data and likelihoods. It may be used to indicate preference of velocities, e.g. peaked around zero like proposed in [6]. The symbol $\sim$ indicates that a proportionality factor normalizing the sum over all distribution elements to $\sum_{\mathbf{v}}\rho^t(\mathbf{v}|\mathbf{x}) = 1$ has to be considered.

If a whole sequence is examined so that previous estimates are available, these estimates of earlier timesteps can be taken into account to update the prior over time, so that instead of Eq. (11) we get the modified Bayes' rule

$$\rho^t(\mathbf{v}|\mathbf{x}) \sim \rho^t(\mathbf{x}|\mathbf{v})\hat{\rho}^t_{\mathbf{x}}(\mathbf{v}) \ . \tag{12}$$

Now the prior $\hat{\rho}^t_{\mathbf{x}}(\mathbf{v})$ is a location $\mathbf{x}$ and time t dependent *prediction* of the velocity pdf at time t using estimates of previous velocity pdfs $\rho^{t-\Delta t}(\mathbf{v}|\mathbf{x})$. This leads to a recursive algorithm essentially consisting of the two stages: 1) *prediction* and 2) *update* of the velocity pdfs $\rho^t(\mathbf{v}|\mathbf{x})$. The update operation uses the latest likelihood $\rho^t(\mathbf{x}|\mathbf{v})$ to calculate the current velocity pdf according to Eq. (12). The prediction operation translates and deforms the previous velocity pdf $\rho^{t-\Delta t}(\mathbf{v}|\mathbf{x})$ to get the new prior $\hat{\rho}^t_{\mathbf{x}}(\mathbf{v})$ for the update operation at measurement time t. This propagation of the velocity pdfs can be calculated in two ways depending on the chosen marginalization approximation:

- Our likelihood distribution $\rho^t(\mathbf{x}|\mathbf{v})$ is calculated with Eq. (10) assuming independently moving image patches so that we neglect correlations based on patch overlaps, with the consequence that the likelihoods at the different locations $\mathbf{x}$ can be seen as being independent from each other. We therefore assume the following *prediction equation for independent patches*:

$$\hat{\rho}^t_{\mathbf{x}}(\mathbf{v}) \sim \rho^{t-\Delta t}(\mathbf{v}|\mathbf{x} - \mathbf{v}\cdot\Delta t) \ . \tag{13}$$

This prediction Eq. (13) is very simple because the probability values of all velocity distributions $\rho^{t-\Delta t}(\mathbf{v}|\mathbf{x})$ of the previous image have to be rearranged assuming only two considerations. First, every pixel moves to a certain degree in every observed direction $\mathbf{v}$. Second, the velocities of all pixels keep constant during time $2\Delta t$ of three consecutive frames. Therefore, linear prediction is applied to all probability values assuming constant velocity $\mathbf{v}^t_{\mathbf{x}} = \mathbf{v}^{t-\Delta t}_{\mathbf{x}-\Delta\mathbf{x}}$. This results in a parallel shift $\Delta\mathbf{x} = \mathbf{v}\Delta t$ in image space $\mathbf{x}$ of all probability values $\rho^{t-\Delta t}(\mathbf{v}|\mathbf{x} - \Delta\mathbf{x})$ that belong to the same velocity $\mathbf{v}$. Afterwards, the shifted values have to be normalized in velocity space to ensure that the sum over all probability values of the predictive prior $\hat{\rho}^t_{\mathbf{x}}(\mathbf{v})$ belonging to one location $\mathbf{x}$ is one $\sum_{\mathbf{v}}\hat{\rho}^t_{\mathbf{x}}(\mathbf{v}) = 1$. In Fig. 1, this leads to simple shift operations of the layers of the velocity distribution map followed by columnwise normalization of the shifted probability values.

- If we want to consider *spatial coherence*, which is known to be important in motion perception [24] and has already been done on the basis of velocity vectors [17], we couple the estimates of velocity pdfs over space in order to get a prior $\hat{\rho}^t_{\mathbf{x}}(\mathbf{v})$ that is dependent on previous estimates of several image locations in a neighborhood $\mathbf{x}'$ around $\mathbf{x}$. The single contributions can be weighted according to the overlap of the patches and the size of the patch window using a weighting window $\hat{\mathbf{W}}$ (we have usually chosen $\hat{\mathbf{W}}$ close to $\mathbf{W}$). This leads to a combined space-time integration that is formulated in the *prediction equation for correlated patches*:

$$\hat{\rho}^t_{\mathbf{x}}(\mathbf{v}) \sim \sum_{\mathbf{x}'} \hat{W}^{\mathbf{x}}_{\mathbf{x}'}\rho^{t-\Delta t}(\mathbf{v}|\mathbf{x}' - \mathbf{v}\cdot\Delta t) \ . \tag{14}$$

This extended prediction scheme that is able to take into consideration neighborhood relations in image space requires an additional correlation of every layer of the

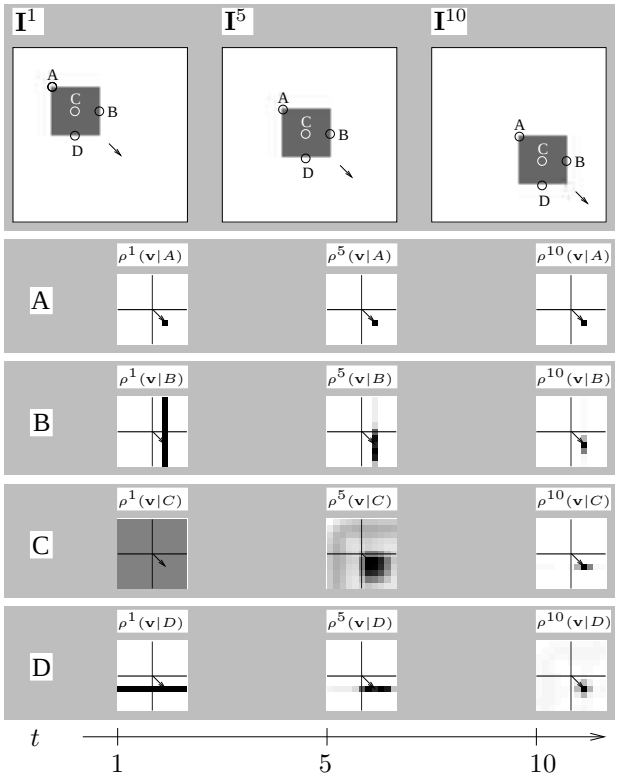

Fig. 3. Example of space-time integration resolving motion ambiguities like the aperture problem at location B and D and the lack of contrast within image regions at location C.

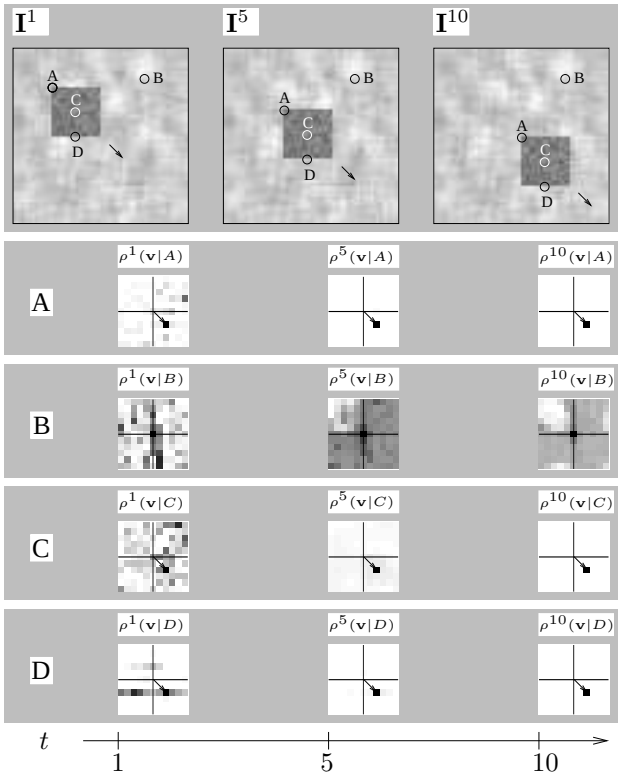

Fig. 4. Example of space-time integration reducing motion ambiguities caused by image noise. Zero mean Gaussian noise with $\sigma_\eta = 20$ is added to the gray values $\mathbf{G} \in [0; 255]$ of the image $\mathbf{I}^t$ at each timestep t.

velocity distribution map with the window $\hat{\mathbf{W}}$. The only assumption is that neighboring locations $\mathbf{x}'$ which are nearer to $\mathbf{x}$ have distributions $\rho^{t-\Delta t}(\mathbf{v}|\mathbf{x}')$ which more closely resemble the distribution $\rho^{t-\Delta t}(\mathbf{v}|\mathbf{x})$ at the observed location $\mathbf{x}$. This implies, it is more likely that these pixels belong to one coherently moving object. In the implementation, we have chosen a Gaussian window so that $\hat{W}^{\mathbf{x}}_{\mathbf{x}'} = e^{-\frac{1}{\sigma^2_{\hat{W}}}||\mathbf{x}-\mathbf{x}'||^2}$.

The prediction Eq. (14) can be seen in analogy to the prediction processes to obtain the prior pdf at the next timestep using the well-known Chapman-Kolmogorov equation (see e.g. [21]), which is the basis for many non-Gaussian Bayesian tracking methods like *p*article filtering [28], [7] or the *Condensation algorithm* [20]. One difference to existing work on the topic of motion estimation is that our prediction includes spatial coherence effects which means correlations between neighboring pixels in velocity space that prefer coherently moving pixels within a neighborhood defined by $\hat{\mathbf{W}}$. In our approach, the prediction of one state does not only depend on the distribution of the previous state and the propagation of its previous distribution. Instead, the propagation of every velocity distribution is dependent on the distributions of all the neighboring pixel $\mathbf{x}'$ that are able to move to the observed location $\mathbf{x}$. Therefore a clear prediction can only be made if the neighboring distributions $\rho^{t-\Delta t}(\mathbf{v}|\mathbf{x}')$ at time $t - \Delta t$, shifted like explained in Eq. (13), are consistent with the momentary observed distribution $\rho^t(\mathbf{v}|\mathbf{x})$ at time t, meaning that all pixels in the neighborhood move homogeneously. The more diverse the distributions are the less certain the prediction is.

In Fig. 3 an example is given how the space-time propagation based on velocity distributions resolves motion ambiguities caused by the aperture problem and the lack of contrast within image regions. As expected, the ambiguities at the edges of the moving square which are reflected in the equally distributed velocity distribution along the moving boundaries are resolved timestep by timestep leading to unimodally distributed clearly peaked velocity distributions. In Fig. 4 another example of a test scene disturbed by noise, e.g. zero mean Gaussian noise added to the image gray values $\mathbf{G} \in [0; 255]$ with variance $\sigma_\eta = 20$, is shown. Again the ambiguities

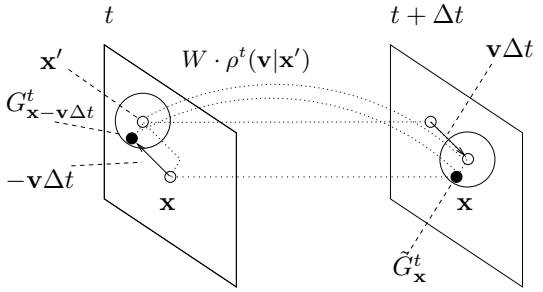

Fig. 5. Principle of warping the image using velocity distributions. Prediction $\tilde{\mathbf{G}}^t$ of the consecutive image $\mathbf{G}^{t+\Delta t}$ using all the velocity distributions $\rho^t(\mathbf{v}|\mathbf{x})$ and all the gray values $G^t_{\mathbf{x}}$ of all overlapping patches in the image $\mathbf{I}^t$ means that each gray value $\tilde{G}^t_{\mathbf{x}}$ is generated by the overlap of all gray values $G^t_{\mathbf{x}-\mathbf{v}\cdot\Delta t}$ of all patches $\mathbf{W}^{\mathbf{x}'}\odot\mathbf{G}^t$ in the neighborhood $\mathbf{x}'$ that overlap the location $\mathbf{x}$ weighted with the weights of the window $\mathbf{W}$ and the probability $\rho^t(\mathbf{v}|\mathbf{x}')$.

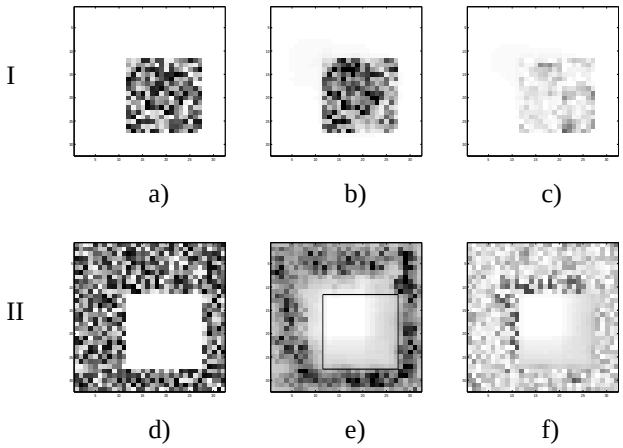

Fig. 6. Two examples of image warping. The first I example is a noisy square moving with four pixel/frame in front of a white background. The results are displayed in the first row of Fig. 6. In a) the original frame, in b) the result of the warped previous frame (that should in the perfect case be identical to a)) and in c) the error between original and warped frame is shown. The second II example shows a white square (d)) moving in front of a noisy background with the same velocity as in I with the warping result e) and the error f). In e) we have marked the outline of the square for clarity.

at the beginning of the sequence are resolved after a few timesteps Δt and the space-time propagation works properly in the case of noisy sequences.

V. Image warping model

Before introducing integration over scale, the needed warping model is proposed and explained. Normally, warping is done by shifting every image gray value $G^t_{\mathbf{x}}$ located at $\mathbf{x}$ to the estimated new position $\mathbf{x}+\Delta\mathbf{x}$ described by the corresponding velocity vector $\mathbf{v}^t_{\mathbf{x}}$ of the optical flow field. In doing so, it can happen that two gray values at different positions in image $\mathbf{I}^t$ are shifted to the very same position in image $\mathbf{I}^{t+\Delta t}$. Therefore some image positions at time $t+\Delta t$ are not filled with corresponding gray values from time t. That means gray value $G^t_{\mathbf{x}}$ to gray value $G^{t+\Delta t}_{\mathbf{x}}$ correspondence over time is not clearly defined over the whole image because of ambiguities in velocity estimation and the appearance or disappearance of gray values at occlusion boundaries. Usually warping is followed by an interpolation step to reduce the errors and fill the "gaps" caused by the just mentioned problems [[18]]. We start with the underlying generative model of locally independent moving image patches for the measurement of local velocity likelihoods $\rho^t(\mathbf{x}|\mathbf{v})$ as formulated in Eq. (2). Now the question is how to formulate a generative model for the whole image warping process reducing the errors made by motion ambiguities without explicit interpolation. To this end, we combine the idea of locally moving and overlapping weighted image $\mathbf{I}^t$ patches $\mathbf{W}\odot\mathbf{G}^{t,\mathbf{x}}$ and the measured local velocity likelihoods $\rho^t(\mathbf{x}|\mathbf{v})$. Again, as before in the propagation over time approach, the likelihoods are thought to be locally dependent according to the overlap of the patches. This implies that $\mathbf{G}^{t+\Delta t}$ is generated by the overlap of all image patches $\mathbf{W}\odot\mathbf{G}^{t,\mathbf{x}}$ moving with all possible velocities and weighted by the posterior velocity distribution $\rho^t(\mathbf{v}|\mathbf{x})$ that is calculated by the likelihoods using Bayes' rule (Eq. (11)). This leads to the following warping process:

$$\tilde{\mathbf{G}}^t := \sum_{\mathbf{v},\mathbf{x}} \rho^t(\mathbf{v}|\mathbf{x})\mathbf{W}^{\mathbf{x}-\mathbf{v}\Delta t}\odot\mathbf{G}^t \ , \qquad (15)$$

with $\tilde{\mathbf{G}}^t$ being the prediction of $\mathbf{G}^{t+\Delta t}$ using all the information ('old' image $\mathbf{G}^t$ and velocity estimation $\rho^t(\mathbf{v}|\mathbf{x})$) available at time t. Figure 5 illustrates the contribution of a pixel of a moving patch anchored at location $\mathbf{x}'$ at time t to the gray value at location $\mathbf{x}$ in image $\mathbf{I}^{t+\Delta t}$. In Fig. 6 two examples of the warping process are shown. The first example is a noisy square moving with four pixel/frame in front of a white background. The results are displayed in the first row I of Fig. 6. In a) the original frame, in b) the result of the warped previous frame and in c) the error between original and warped frame is shown. In this example the background is clutter-free and therefore no errors caused by appearing and disappearing gray values can occur. The errors result from the ambiguity of the velocity distributions and are dependent on the spread of the distributions. The extreme case of single peak distributions would cause no errors at all resulting in a perfect prediction by the warping process. The more the distributions are spread the more blurred are the warped images. The second row II shows a second example of a white square moving in front of a noisy background with the same velocity as in I. Now

gray values appear and disappear over time. Therefore correlation results of corresponding patches are less unambiguous at occlusion boundaries and the spread and ambiguity of these velocity distributions increases. That leads to blurred object edges (e)) and has an effect similar to the interpolation and consequent smoothing/low-pass filtering used in standard warping methods.

VI. Integration over Scales

Now we regard a coarse-to-fine hierarchy of velocity detectors [29]. A single level of the hierarchy is determined by 1) the resolution of the images that are processed 2) the range of velocities that are scanned and 3) the window $\mathbf{W}$ of the patches that are compared. Coarser spatial resolutions correlate with higher velocities and larger patch windows. The strategy proceeds from coarse to fine, i.e., first the larger velocities are calculated, then smaller relative velocities, then even smaller ones, etc.

In the resolution pyramid, at each level k we have different velocity distributions $\rho_k^t(\mathbf{v}|\mathbf{x})$ for the same absolute velocity $\mathbf{v}$ at its corresponding image location $\mathbf{x}$. The calculation of the pdfs includes space-time propagation on each level in the manner as introduced in section IV. Velocity pdfs at higher levels of the pyramid (i.e., using lower spatial resolutions) are calculated using larger windows $\mathbf{W}$, therefore showing a tendency towards less aperture depending problems but more estimation and integration errors. To the contrary, velocity pdfs at lower levels of the pyramid (higher resolutions) tend to be more accurate but also more prone to aperture problems.

Nevertheless, the velocity pdfs at the different levels of the pyramid are not independent of each other. The purpose of the pyramid is therefore to couple the different levels of velocity pdfs in order to 1) gain a coarse-to-fine description of velocity estimations 2) take advantage of more global pdfs to reduce motion ambiguities and 3) use the more local pdfs to gain a highly resolved velocity signal. The goal is to be able to simultaneously estimate high velocities yet retain fine velocity discrimination abilities.

In order to achieve this, we do the following: The highest level of the pyramid estimates large-scale/large-region velocity pdfs of the image. These velocity pdfs are used to impose a moving reference frame for the next lower pyramid level to estimate better resolved, more local velocity pdfs. That is, we decompose the velocity distributions in a coarse-to-fine manner, estimating at each level the relative velocity distributions needed for an accurate total velocity distribution estimation.

There are several advantages of such a procedure. If we want to get good estimates for both large and highly resolved velocities/distributions without a pyramidal structure, we would have to perform calculations for each possible velocity, which is computationally prohibitive. In a pyramidal structure, we get increasingly refined estimations for the velocities starting from inexpensive, but coarse initial approximations and refining further at every level.

At each level of the pyramid, we do the following calculations:

- Start with the gray value inputs

$$\tilde{\mathbf{G}}_k^t,\ \mathbf{G}_k^{t+\Delta t}\ . \tag{16}$$

$\tilde{\mathbf{G}}_k^t$ is the level k prediction of all gray values of image $\mathbf{I}_k^{t+\Delta t}$, using the information available at t, calculated with our warping model introduced in section V. With $k=0$ denoting the highest level we have $\tilde{\mathbf{G}}_0^t = \mathbf{G}_0^t$ (i.e. the first estimation is directly given by the measurement), since there are no further assumptions about velocities $\mathbf{v}$.

- Calculate the local likelihood for the k-th level velocity $\tilde{\mathbf{v}}$

$$\tilde{\rho}_k^t(\mathbf{x}|\tilde{\mathbf{v}}) \sim e^{-\frac{1}{2}\cdot\left(\frac{\sigma_{\tilde{\mathbf{G}}_k^{t,\mathbf{x}}}}{\sigma_\eta}\right)^2\left(1-\varrho^2_{\tilde{\mathbf{G}}_k^{t,\mathbf{x}},\mathbf{G}_k^{t+\Delta t,\mathbf{x}+\tilde{\mathbf{v}}\Delta t}}\right)} \tag{17}$$

as formulated in Eq. (10). Note that at the highest level, $\tilde{\mathbf{v}}$ is equal to the *absolute velocity* $\mathbf{v}$ from $\rho_k^t(\mathbf{x}|\mathbf{v})$, whereas at lower levels, $\tilde{\mathbf{v}}$ is a *differential/relative* velocity related with the likelihood $\tilde{\rho}_k^t(\mathbf{x}|\tilde{\mathbf{v}})$. Note also that $\tilde{\rho}_k^t(\mathbf{x}|\tilde{\mathbf{v}})$ correlates $\tilde{\mathbf{G}}_k^{t,\mathbf{x}}$ (and not $\mathbf{G}_k^{t,\mathbf{x}}$) with $\mathbf{G}_k^{t+\Delta t,\mathbf{x}+\tilde{\mathbf{v}}\Delta t}$.

- Calculate the local likelihood $\rho_k^t(\mathbf{x}|\mathbf{v})$ for the absolute velocity $\mathbf{v}$ by combining the posterior estimation $\rho_{k-1}^t(\mathbf{v}|\mathbf{x})$ for the absolute velocity $\mathbf{v}$ from the higher stage $k-1$ with the likelihood estimations for the relative velocity $\tilde{\mathbf{v}}$ from stage k as follows:

$$\begin{gathered}\rho_k^t(\mathbf{x}|\mathbf{v}) :\sim \\ \textstyle\sum_{\tilde{\mathbf{v}}}\sum_{\tilde{\mathbf{v}}=\mathbf{v}-\check{\mathbf{v}}}\tilde{\rho}_k^t(\mathbf{x}+(\mathbf{v}-\tilde{\mathbf{v}})\Delta t|\tilde{\mathbf{v}})\,\check{\rho}_k^t(\check{\mathbf{v}}|\mathbf{x})\,, \\ \text{with}\qquad \check{\rho}_k^t = \mathcal{I}(\rho_{k-1}^t(\mathbf{v}|\mathbf{x}))\,.\end{gathered} \tag{18}$$

Function $\mathcal{I}$ interpolates the estimate $\rho_{k-1}^t(\mathbf{v}|\mathbf{x})$ from coarser scale to resolution of scale k. At the highest level there will be no combination because no velocity distributions from a coarser level are available and therefore $\rho_0^t(\mathbf{x}|\mathbf{v}) = \tilde{\rho}_0^t(\mathbf{x}|\mathbf{v})$ because there we are directly considering the absolute velocity.

- Combine the likelihood $\rho_k^t(\mathbf{x}|\mathbf{v})$ with the prior $\hat{\rho}_{\mathbf{x},k}^t(\mathbf{v})$ gained from **space-time propagation** (Eq. (14))

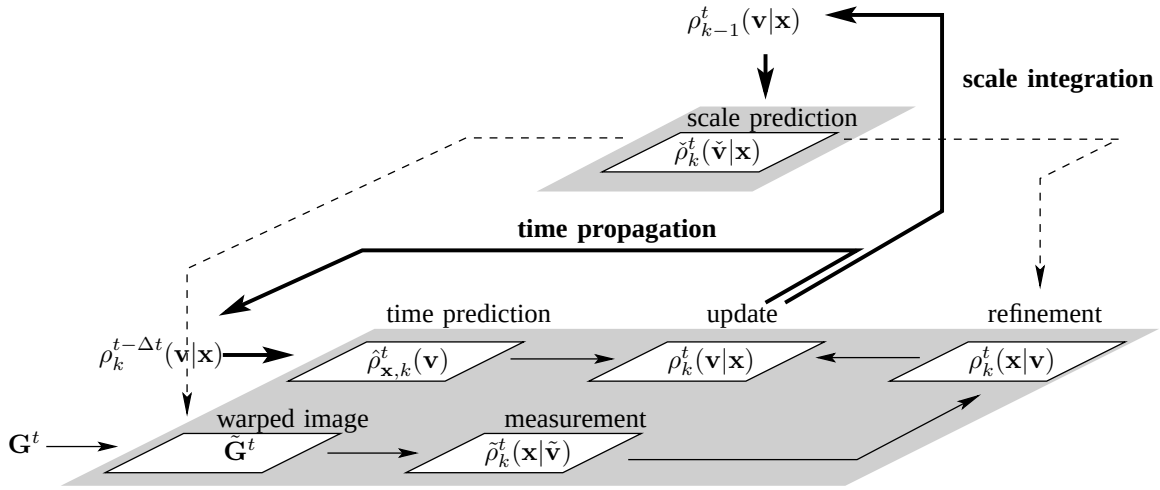

Fig. 7. Block diagram of the estimation process over space-time and scales. The momentary estimation $\rho^t_k(\mathbf{v}|\mathbf{x})$ combines the prediction from coarser scale $\check{\rho}^t_k(\check{\mathbf{v}}|\mathbf{x})$ with the prediction from previous time $\hat{\rho}^t_{\mathbf{x},k}(\mathbf{v})$ and the local measurement $\tilde{\rho}^t_k(\mathbf{x}|\tilde{\mathbf{v}})$. The scale integration of the velocity distribution is calculated at every time t and space-time propagation is done for every hierarchy in parallel within the time interval $[t, t+\Delta t]$.

at level k using the posterior distributions $\rho^{t-\Delta t}_k(\mathbf{v}|\mathbf{x})$ at time $t-\Delta t$ to get the posterior distribution $\rho^t_k(\mathbf{v}|\mathbf{x})$ for the absolute velocity $\mathbf{v}$ according to

$$\rho^t_k(\mathbf{v}|\mathbf{x}) \sim \rho^t_k(\mathbf{x}|\mathbf{v})\,\hat{\rho}^t_{\mathbf{x},k}(\mathbf{v})\ . \tag{19}$$

• Use the gained posterior distribution for the warping of the image $\mathbf{G}^t_{k+1}$ to time $t+\Delta t$ at the next level $k+1$ resulting in a predictive image $\tilde{\mathbf{G}}^t_{k+1}$ according to

$$\tilde{\mathbf{G}}^t_{k+1} := \sum_{\mathbf{v},\mathbf{x}} \rho^t_k(\mathbf{v}|\mathbf{x})\mathbf{W}^{\mathbf{x}-\mathbf{v}\Delta t} \odot \mathbf{G}^t_{k+1}\ . \tag{20}$$

This is the best estimate according to level k, time t and the constraints given by the warping model.

• Increase the pyramid level k and repeat the procedure.

The block diagram of the iteration process is shown in Fig. 7. Examples of the reduction of motion ambiguities through scales are shown in Fig. 8. With increasing scale level k the velocity distributions $\rho^t_k(\mathbf{v}|\mathbf{x})$ at location $\mathbf{x} = A$ and $\mathbf{x} = B$ are refined. Especially multiple peaked distributions change to more or less unimodal distributions.

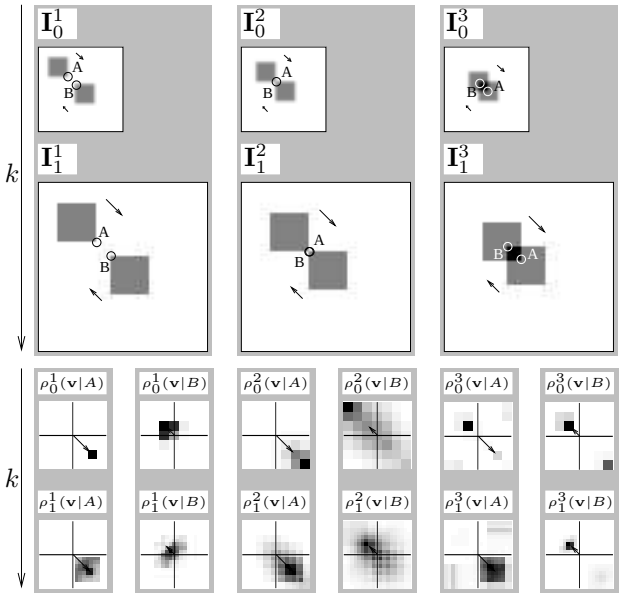

Fig. 8. Example of scale integration reducing motion ambiguities caused by multiple motions and representing high velocities at high resolution in velocity space.

VII. Results

In general, for the presented method it is not necessary to optimize the parameters and interpolation methods according to the object sizes in the scene and the underlying real movement patterns. Only the noise parameter σ_η, the size of the patches $\mathbf{W}$, the integration window $\hat{\mathbf{W}}$ and the search area for the matching algorithm have to be chosen. No informations about object sizes and movement characteristics are incorporated. Moreover, the target is to find a tradeoff between accuracy of the estimated fbw field and computational cost.

To achieve this, the maximum velocity that the system should detect has to be defined. According to this velocity the number of discrete probability values per distribution has to be chosen. The patch size determined

A Yosemite fbw B Yosemite probabilities C Translating Tree fbw

Fig. 9. Examples of (A) the flow field of the Yosemite sequence with its (B) corresponding probability values (the whiter the pixels are the more probable the velocity is) and (C) the flow field of the Translating Tree sequence often used as benchmark for optical flow computation. With our space-time-scale integration framework applied over 9 consecutive frames we achieve a mean angular error of 8.51° on the cloudy Yosemite sequence and a mean angular error of 0.34° on the Translating Tree sequence.

by the size of window $\mathbf{W}$ is normally set in the range of 3×3 up to 9×9 pixels. The weighting of the patches $\mathbf{W}$ as well as of the neighborhood $\hat{\mathbf{W}}$ was done using a symmetrical Gaussian window with variances $\sigma_{\mathbf{W}}$,$\sigma_{\hat{\mathbf{W}}}$ half the size of the window. The noise parameter $\sigma_\eta = \alpha\bar{\sigma}_{\mathbf{G}^{t,\mathbf{x}}}$ can be tuned, e.g. using the mean variance $\bar{\sigma}_{\mathbf{G}^{t,\mathbf{x}}}$ of all local variances of the momentary input image $\sigma_{\mathbf{G}^{t,\mathbf{x}}}$ scaled with a factor α (we have chosen the same $\alpha = 0.5$ for all calculations). For the space-time-scale algorithm the image decomposition is done with a standard Gaussian pyramid [25]. The initial priors $\hat{\rho}_k^0(\mathbf{v}|\mathbf{x})$ are chosen Gaussian distributed with variance 0.25 times the maximum velocity, which means preferring zero velocities at the beginning. For all examples where the fbw field was calculated the MMSE estimator [30] is used for fbw field estimation.

$$\mathbf{v}_{\text{MMSE}} = \sum_{\mathbf{v}} \rho_k^t(\mathbf{v}|\mathbf{x}) \tag{21}$$

Evaluations to judge the accuracy, the robustness against noise and parameter variations as well as the computation time have been done on the Yosemite sequence [31] with and without cloudy sky created by Lynn Quam and the Translating Tree sequence [31] created by David Fleet (see Fig. 9). Although our paper does not focus on high accuracy, our technique compares quite favorably to the techniques reported in [31]. In Table I the best techniques listed in [31] are shown including our results. The used error statistics that combine angular and magnitude error denoted as AAE = *average angular error* and STD = *standard deviation* are also proposed in [31]. The most convincing result for our technique is the AAE of 0.34° for the Translating Tree sequence. But also for the Yosemite sequence our results are better than all the results reported in [31] even with additive Gaussian noise with standard deviation of $\sigma_G = 40$ as listed in Table II.

Best results on the cloudy Yosemite sequence reported in [31]			
Technique	AAE	STD	Density
Anandan	15.84°	13.46°	100%
Singh	13.16°	12.07°	100%
Nagel	11.71°	10.59°	100%
Horn and Schunk mod.	11.26°	16.41°	100%
Uras et al.	10.44°	15.00°	100%
Fleet and Jepson	4.29°	11.24°	34.1%
Lukas and Kanade	3.05°	7.31°	8.7%
Our results on the cloudy Yosemite sequence (see also Fig. 9)			
Int. over space-time	8.73°	10.45°	100%
Int. over space-time-scale	8.51°	10.62°	100%
	2.88°	5.02°	34%
Our results on the cloudless Yosemite sequence			
Int. over space-time	6.29°	6.56°	100%
Best results on the Translating Tree sequence reported in [31]			
Nagel	2.44°	3.06°	100%
Singh	1.25°	3.29°	100%
Uras et al.	0.62°	0.52°	100%
Our results on the Translating Tree sequence (see also Fig. 9)			
Int. over space-time-scale	0.34°	0.12°	100%

TABLE I

BENCHMARK REPORTED IN [31] INCLUDING OUR RESULTS FOR COMPARISON (AAE = AVERAGE ANGULAR ERROR, STD = STANDARD DEVIATION).

The probability values (see B in Fig. 9) serve as a good confidence measure to exclude wrong estimates and reduce the density of the fbw field but increase the correctness. The robustness against parameter variations on the patch window size $\mathbf{W}$ and the integration window size $\hat{\mathbf{W}}$ are shown in Table III. In Table IV some measurements on the computation time have been done for different image sizes and velocity search spaces.

The computations have been performed on a 1.8 GHz Intel Pentium 4 processor executing C code. Since we are using a correlation-based method the number of

Fig. 10. Example of (A) the flow field of the Darmstadt Traffic sequence, (B) the corresponding probability values and (C) the magnitude of the extracted velocity vectors.

operations for a single scale-level is in the order of $imagesize \times patchsize \times number\,of\,velocities$. But nearly all formulations can be implemented very efficiently using accelerated correlation algorithms because our algorithm offers the possibility for a highly parallel computation. In particular, we used the Fast Fourier Transform for all the correlational computations.

In Fig. 10 we give an example of a real-world sequence including camera noise and changing lightning conditions that is more difficult to treat than the used benchmark sequences. The first image (A) shows the optical flow, in (B) the corresponding probability values are plotted and in (C) the magnitude of the extracted velocity vectors are shown. We want to mention that there are some gradient based techniques reported in [16] focusing on high accuracy optical flow computation using a continuous, rotationally invariant energy functional with a non-linearized data term. With this energy functional that is minimized iteratively and depends on several parameters even better results can be achieved. Our estimates are based on a simple linear generative model and do not have to be calculated iteratively within a timestep but they improve from frame to frame with integration over space, time and scale.

σ_G	0	10	20	30	40
AAE	8.50°	9.33°	9.62°	10.01°	10.38°
STD	10.60°	11.09°	11.54°	12.02°	12.53°

TABLE II

RESULTS FOR THE ACCURACY DECREASE WHEN GAUSSIAN NOISE WITH INCREASING STANDARD DEVIATION σ_G IS ADDED TO THE CLOUDY YOSEMITE SEQUENCE.

$\mathbf{W}$	5×5	5×5	5×5	3×3	7×7	9×9
$\hat{\mathbf{W}}$	5×5	15×15	35×35	35×35	35×35	35×35
AAE	7.58°	7.06°	6.40°	6.76°	**6.29°**	6.45°
STD	8.35°	7.36°	6.42°	6.40°	**6.56°**	6.84°

TABLE III

RESULTS FOR PARAMETER VARIATIONS ON THE PATCH WINDOW $\mathbf{W}$ AND THE SPATIAL CORRELATION WINDOW $\hat{\mathbf{W}}$ ON THE CLOUDLESS YOSEMITE SEQUENCE.

image size [pixel × pixel]	128×128	128×128
sampled $\mathbf{v}$ [pixel/framerate]	25 pix/fr	81 pix/fr
comp. time [seconds/frame]	0.09 sec/f	0.28 sec/f

256×256	256×256	256×256	252×316 Yosemite
9 pix/fr	25 pix/fr	81 pix/fr	81 pix/fr
0.24 sec/f	0.61 sec/f	1.48 sec/f	1.67 sec/f

TABLE IV

COMPUTATION TIME FOR DIFFERENT DATA SIZES.

VIII. CONCLUSION

In this paper we have presented a motion estimation framework based on a distributed representation of velocity information. We have extended known concepts of optical flow estimation dealing with velocity vectors towards non-Gaussian velocity distributions, introducing brightness and contrast invariance as well as propagation of motion information over space, time and scales. Our joint integration model is able to resolve motion ambiguities and remains robust in the case of image noise and brightness variation. It proposes a solution to the aperture and the blank wall problem and is applicable to real-world sequences. The implemented algorithm is surprisingly fast without using specific hardware and the system architecture can be interpreted in a more biological fashion. It can be used as a basic mid-level motion processing module to build upon higher-level systems, like motion segmentation and multiple object tracking.

In general our approach is constructed to resolve ambiguities in coherently moving regions. Additional a priori knowledge about object boundaries respectively

segmentation information, gained from static features like e.g. edges, could be explicitly considered in the spatial integration of motion information and should improve the results and sharpen the overall pdfs.

REFERENCES

[1] B. K. P. Horn and B. G. Schunk, "Determining optic flow," *Artificial Intelligence*, vol. 17, pp. 185–204, 1981.

[2] S. Beauchemin and J. Barron, "The computation of optical flow," *ACM Computing Surveys*, vol. 27(3), pp. 433–467, 1995.

[3] 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*, vol. 10(1), pp. 56–99, 2001.

[4] E. Simoncelli, E. Adelson, and D. Heeger, "Probability distributions of optical flow," in *IEEE CVPR*, pp. 310–315, 1991.

[5] E. Simoncelli, "Distributed representation and analysis of visual motion," *PhD thesis, MIT Department of Electrical Engineering and Computer Science*, 1993.

[6] Y. Weiss, "Bayesian motion estimation and segmentation," *PhD thesis, MIT Department of Brain and Cognitive Science*, 1998.

[7] J. Zelek, "Bayesian real-time optical flow," in *15th Int. Conf. on Vision Interface*, pp. 310–315, 2002.

[8] A. Singh, "An estimation-theoretic framework for image-flow computation," in *3rd IEEE ICCV*, pp. 168–177, 1990.

[9] Q. X. Wu, "A correlation-relaxation-labeling framework for computing optical flow - template matching from a new perspective," *IEEE Trans. PAMI*, vol. 17(9), pp. 843–853, 1995.

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

[11] B. D. Lukas and T. Kanade, "An iterative image-registration technique with an application to stereo vision," *DARPA Image Understanding Workshop*, pp. 121–130, 1981.

[12] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *IJCV*, vol. 2, pp. 283–310, 1989.

[13] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *2nd ECCV*, pp. 237–252, 1992.

[14] J. Weber and J. Malik, "Robust computation of optical flow in a multi-scale differential framework," *IJVC*, vol. 14(1), pp. 5–19, 1995.

[15] E. Memin and P. Perez, "A multigrid approach for hierarchical motion estimation," *6th ICCV*, pp. 933–938, 1998.

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

[17] A. Singh, "Incremental estimation of image flow using a kalman filter," in *IEEE Workshop on Visual Motion*, pp. 36–43, 1991.

[18] B. Jähne, H. Haussecker, and P. Geissler, eds., *Handbook of Computer Vision and Applications*, ch. Bayesian Multi-Scale Differential Optical Flow, pp. 397–421. Academic Press, 1999.

[19] Y. Rosenberg and M. Werman, "Representing local motion as a probability distribution matrix applied to object tracking," in *IEEE CVPR*, pp. 654–659, 1997.

[20] M. Isard and A. Blake, "Condensation - conditional density propagation for visual tracking," *IJCV*, vol. 29, pp. 5–28, 1998.

[21] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line nonlinear/non-gaussian bayesian tracking," *IEEE Trans. on Signal Processing*, vol. 50(2), pp. 174–188, 2002.

[22] S. Negahdaripour, "Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis.," *IEEE Trans. PAMI*, vol. 20(9), pp. 961–979, 1998.

[23] H. Haussecker and D. Fleet, "Computing optical flow with physical models of brightness variation," *IEEE CVPR*, vol. 5, pp. 77–104, 1990.

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

[25] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. on Communications*, vol. COM-31,40, pp. 532–540, 1983.

[26] O. Nestares, D. Fleet, and D. Heeger, "Likelihood functions and confidence bounds for total-least-squares problems," *IEEE CVPR*, vol. 2, pp. 760–767, 2000.

[27] R. Duda, P. Hart, and D. Stork, *Pattern Classification,2nd Edition*. John Wiley and Sons(UK) LTD, Chichester, 2001.

[28] Y. Rosenberg and M. Werman, "A general filter for measurements with any probability distribution," in *IEEE CVPR*, pp. 654–659, 1997.

[29] J. Eggert, V. Willert, and E. Körner, "Building a motion resolution pyramid by combining velocity distributions," *26th DAGM Symposium*, vol. LNCS 3175, pp. 310–317, 2004.

[30] J. Bernardo and A. Smith, *Bayesian Theory*. No. 5 in Wiley Series in Probability and Statistics, John Wiley & Sons (UK) LTD, Chichester, 2004.

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