Building a Motion Resolution Pyramid by Combining Velocity Distributions

Julian Eggert, Volker Willert, Edgar Körner

2004

Preprint:

This is an accepted article published in 26th Pattern Recognition Symposium DAGM. The final authenticated version is available online at: https://doi.org/[DOI not available]

Proceedings of the Pattern Recognition Symposium DAGM 2004, pp. 310-317

Building a Motion Resolution Pyramid by Combining Velocity Distributions

Julian Eggert[1], Volker Willert[2], and Edgar Körner[1]

[1] HRI Honda Research Institute GmbH,
Carl-Legien-Straße 30, 63073 Offenbach/Main
{Julian.Eggert, Edgar.Koerner}@honda-ri.de
[2] TU Darmstadt, Institut für Automatisierungstechnik
Fachgebiet Regelungstheorie & Robotik,
Landgraf-Georg-Str.04, 64283 Darmstadt
volker@rtr.tu-darmstadt.de

Abstract. Velocity distributions are an enhanced representation of image velocity implying more velocity information than velocity vectors. Velocity distributions allow the representation of ambiguous motion information caused by the aperture problem or multiple motions at a given image region. Starting from a contrast- and brightness-invariant generative model for image formation a likelihood measure for local image velocities is proposed. These local velocities are combined into a coarse-to-fine-strategy using a pyramidal image velocity representation. On each pyramid level, the strategy calculates predictions for image formation and combines velocity distributions over scales to get a hierarchically arranged motion information with different resolution levels in velocity space. The strategy helps to overcome ambiguous motion information present at fine scales by integrating information from coarser scales. In addition, it is able to combine motion information over scales to get velocity estimates with high resolution.

1 Introduction

Traditionally, motion estimates in an image sequence are represented using vector fields consisting of velocity vectors each describing the motion at a particular image region or pixel. Yet in most cases single velocity vectors at each image location are a very impoverished representation, which may introduce great errors in subsequent motion estimations. This may, e.g., be because the motion measurement process is ambiguous and disturbed by noise. The main problems which cause these errors are the aperture problem, the lack of contrast within image regions, occlusions at motion boundaries and multiple motions at local image regions caused by large image regions or transparent motion.
To circumvent these problems, the velocity of an image patch at each location is understood as a statistical signal. This implies working with probabilities for the existence of image features like pixel gray values and velocities. The expectation is that probability density functions are finally able to tackle the addressed

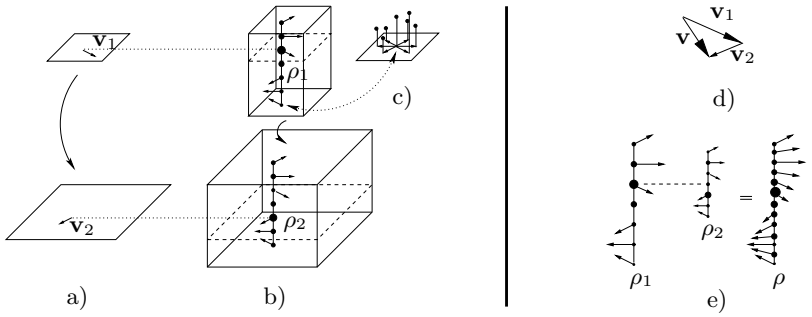

Fig. 1. Comparison of multiscale image motion representation: a) velocity vectors $\mathbf{v}_1, \mathbf{v}_2$ as standard representation and b) velocity distributions ρ_1, ρ_2 as enhanced representation. The single column with arrows in b) represents a velocity distribution at one location; it is shown in c) as the corresponding 2-dimensional graph. In d) and e) the velocity decomposition principle is shown schematically. In d), we assume that the true velocity $\mathbf{v}$ is decomposed into velocity vectors (e.g. $\mathbf{v}_1$ and $\mathbf{v}_2$ in the figure) at different scales. In e), we do the analog procedure for velocity distributions: We combine $\rho_1(\mathbf{v}_1)$ and $\rho_2(\mathbf{v}_2)$ in such a way that we get a total $\rho(\mathbf{v})$ with $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$.

problems related to motion processing, like ambiguous motion, occlusion and transparency, since some specific information about them can in principle be extracted from the probability functions [3]. During the last ten years velocity distributions have been motivated by several authors [3], [4], [5] mainly using two approaches: the gradient based *brightness change constraint equation* and the correlation-based *patch matching* technique.

A problem when dealing with velocity distributions is how to represent them in a multiscale pyramid. Such a pyramid 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 highest *velocities* (connected with the coarsest spatial resolution) are calculated first, then a shifted [1] version of the image is calculated using the velocities, and afterwards the velocities of the next pyramid level are calculated. These then correspond to *relative* velocities because they have been calculated in a frame that is moving along with the velocities from the coarse resolution. But still, single velocities are used for the shifted version of the image, so that the information available in the *distribution* is neglected.

In this work, we first introduce a *linear generative model* of image patch formation over time. Here, the changes in two consecutive images depend on the displacements as well as brightness and contrast variations (see Eq.2) of localized image patches. The result are contrast and brightness invariant velocity distributions based on a correlation measure comparing windowed image patches of consecutive images. Afterwards, we set up a hierarchical chain of velocity distributions from coarse to fine spatial scale and from large to smaller (relative) velocities. At each stage of the pyramid, the distributions for the overall velocities are improved using the distributions from the coarser spatial scale as

a starting point and combining it with the local measurements for the relative velocity at the given scale. 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 [1], [2]. The idea on how to combine distributions among scales is illustrated in Fig. 1. The presented architecture combines the advantages of a hierarchical structure and the representation of velocities using distributions and allows for a coarse-to-fine estimation of velocity distributions.

2 Velocity Probability Distributions

In an image sequence[1], 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 $G^t_{\mathbf{x}}$ (scalar) and its velocity vector $\mathbf{v}^t_{\mathbf{x}}$, whereas $\mathbf{G}^t$ denotes the matrix of all gray values of image $\mathbf{I}^t$. The *motion field* is the set of all physical velocities at the corresponding pixel locations at a time t. The *optical flow* is an estimate for the real image motion field at a particular 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 now to assume that all gray values inside of a patch around $\mathbf{x}$ move with a certain common velocity $\mathbf{v}_{\mathbf{x}}$ 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 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 motion estimate more precisely, we recur to a generative model. Our Ansatz is that an image $\mathbf{I}^{t+\Delta t}$ is causally linked with its preceding image $\mathbf{I}^t$ 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 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 (both considered to be constant within a patch) so that we arrive at

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

[1] Notation: We use simple font for scalars (a, A), bold for vectors and matrices $(\mathbf{a}, \mathbf{A})$, and calligrafic font for functions and operators $(\mathcal{A})$. $\mathbf{1}, \mathbf{1}$ are vector of ones and matrix of ones, $\mathbf{A}\odot\mathbf{B}$ denotes a componentwise multiplication of two vectors or matrices and $\mathbf{A}^{@}$ a componentwise exponentiation by α of a vector or matrix.

Assuming that the image noise is zero mean gaussian with variance σ_η, 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 $\mathbf{v}^t_{\mathbf{x}}$, the window function $\mathbf{W}$ and the parameters λ, κ and σ_η, can be written down as[2]:

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

We now proceed to make it less influencial of λ and κ, that means a match of the patches will be almost contrast and brightness invariant. For this purpose, we maximize the likelihood Eq. 3 with respect to the scaling and shift parameters. This amounts to minimizing the exponent, so that we want to find

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

The final result of this minimization process is formulated in Eq. 7. Consider

$$\{\lambda^*,\kappa^*\} := \mathrm{argmin}_{\lambda,\kappa}\left\|\mathbf{W}\odot\left(\lambda\,\mathbf{A}+\kappa\,\mathbf{1}-\mathbf{B}\right)\right\|^2 \quad . \tag{5}$$

$$\text{This leads to } \lambda^* = \frac{\varrho_{\mathbf{A},\mathbf{B}}\cdot\sigma_{\mathbf{B}}}{\sigma_{\mathbf{A}}} \text{ and } \kappa^* = \mu_{\mathbf{B}} - \lambda^*\cdot\mu_{\mathbf{A}} \quad .^3 \tag{6}$$

Inserting Eq. 6 into Eq. 3, so that $\lambda \to \lambda^*$ and $\kappa \to \kappa^*$, leads to the following likelihood formulation

$$\rho^t(\mathbf{x}|\mathbf{v}) := \rho_{\lambda^*,\kappa^*,\sigma_\eta}(\mathbf{G}^{t+\Delta t,\mathbf{x}+\Delta\mathbf{x}}|\mathbf{v}^t_{\mathbf{x}},\mathbf{W},\mathbf{G}^{t,\mathbf{x}}) \sim e^{-\frac{1}{2}\cdot\left(\frac{\sigma_{\mathbf{G}^{t,\mathbf{x}}}}{\sigma_\eta}\right)^2\left(1-\varrho^2_{\mathbf{G}^{t,\mathbf{x}},\mathbf{G}^{t+\Delta t,\mathbf{x}+\Delta\mathbf{x}}}\right)} \quad . \tag{7}$$

The weighted empirical correlation coefficient $\varrho_{\mathbf{A},\mathbf{B}}$ is well known in statistics as an effective template matching measurement. Eq. 7 shows some additional properties according to comparable likelihood measures [4], [3], [5]. The measure $\varrho^2_{\mathbf{G}^{t,\mathbf{x}},\mathbf{G}^{t+\Delta t,\mathbf{x}+\Delta\mathbf{x}}}$ ensures that the match is less affected by local changes in contrast and brightness. Local changes in illumination due to movement of an object when there is a fixed light-source or changes in illumination because of movement of the light-source itself does less reduce the accuracy of the measurement of the likelihood.

Another property of Eq. 7 is given by the ratio of variance of the patch at location $\mathbf{x}$ to the variance of the gaussian distributed noise $\sigma_{\mathbf{G}^{t,\mathbf{x}}}/\sigma_\eta$. The higher this ratio the smaller the overall variance $\sigma = \sigma_\eta/\sigma_{\mathbf{G}^{t,\mathbf{x}}}$. That means, if $\sigma_{\mathbf{G}^{t,\mathbf{x}}}$ is high, then σ is low and mainly the good patch matches contribute to the distribution and it will be clearly peaked. When there is a patch with low variance the distribution will be broader. For higher/lower noise level σ_η, more/less high contrast patches are needed to get a significantly peaked distribution, so that for low σ_η the more also poorly matching results contribute to the likelihood distribution. Therefore σ_η can act as a parameter to control the influence of the variance $\sigma_{\mathbf{G}^{t,\mathbf{x}}}$ of the patch on the confidence of the distribution.

[2] The symbol $\sim$ indicates that a proportional factor normalizing the sum over all distribution elements to 1 has to be considered.

[3] With $\mu_{\mathbf{A}} = \langle\mathbf{A}\rangle := \frac{\mathbf{1}^T\mathbf{A}\odot\mathbf{W}^{②}\mathbf{1}}{\mathbf{1}^T\mathbf{W}^{②}\mathbf{1}}$ and $\sigma^2_{\mathbf{A}} = \langle\mathbf{A}^{②}\rangle - \langle\mathbf{A}\rangle^{②}$, analogous for $\mu_{\mathbf{B}}$ and $\sigma^2_{\mathbf{B}}$, and $\varrho_{\mathbf{A},\mathbf{B}} = \frac{1}{\sigma_{\mathbf{A}}\cdot\sigma_{\mathbf{B}}}\big\langle(\mathbf{A}-\mu_{\mathbf{A}}\mathbf{1})\odot(\mathbf{B}-\mu_{\mathbf{B}}\mathbf{1})\big\rangle = \frac{1}{\sigma_{\mathbf{A}}\cdot\sigma_{\mathbf{B}}}\Big(\langle\mathbf{A}\odot\mathbf{B}\rangle - \mu_{\mathbf{A}}\mu_{\mathbf{B}}\Big)$ being the correlation measure.

3 Coarse-to-fine Strategy

Now we regard a coarse-to-fine hierarchy of velocity detectors. A single level of the hierarchy is determined by (i) the resolution of the images that are compared (ii) the range of velocities that are scanned and (iii) 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.
For a single level of resolution, we use the local velocity estimation

$$\rho^t(\mathbf{v}|\mathbf{x}) \sim \rho^t(\mathbf{x}|\mathbf{v})\rho(\mathbf{v}) \tag{8}$$

with a common prior velocity distribution $\rho(\mathbf{v})$ for all positions $\mathbf{x}$. The prior $\rho(\mathbf{v})$ may be used to indicate preference of velocities, e.g. peaked around zero.
In the resolution pyramid, at each level k we have a different velocity estimation $\rho_k^t(\mathbf{v}|\mathbf{x})$ for the same physical velocity $\mathbf{v}$ at its corresponding physical location $\mathbf{x}$. Velocity estimations 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 errors. To the contrary, velocity estimations at lower levels of the pyramid (higher resolutions) tend to be more accurate but also more prone to aperture problems.
Nevertheless, the estimations at the different levels of the pyramid are not independent of each other. The goal of the pyramid is therefore to couple the different levels in order to (i) gain a coarse-to-fine description of velocity estimations (ii) take advantage of more global estimations to reduce the aperture problem and (iii) use the more local estimations 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 global velocities of the image. These velocities are used to impose a moving reference frame for the next lower pyramid level to estimate better resolved, more local velocities. 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.
The advantages of such a procedure are manifold. 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 steps:

1. Start with inputs
$$\mathbf{G}_k^{t+\Delta t} \ , \ \tilde{\mathbf{G}}_k^t \ . \tag{9}$$
$\tilde{\mathbf{G}}_k^t$ is the level k prediction of all gray values of image $\mathbf{I}_k^{t+\Delta t}$, using the information available at t. E.g., for the highest level with $k = 0$, $\tilde{\mathbf{G}}_0^t = \mathbf{G}_0^{t+\Delta t}$, since there are no further assumptions about velocities $\mathbf{v}$ (i.e. $\mathbf{v} = \mathbf{0}$).
2. Calculate the local likelihood for the k-th level velocity
$$\tilde{\rho}_k^t(\mathbf{x}|\mathbf{v}_k) \sim e^{-\frac{1}{2}\cdot\left(\frac{\sigma_{\tilde{\mathbf{G}}_k^{t,\mathbf{x}}}}{\sigma_\eta}\right)^2\left(1-\varrho^2_{\tilde{\mathbf{G}}_k^{t,\mathbf{x}},\mathbf{G}_k^{t+\Delta t,\mathbf{x}+\Delta\mathbf{x}}}\right)} \tag{10}$$
as formulated in Eq. 7. Note that at the highest level, $\mathbf{v}_0$ is equal to the physical velocity $\mathbf{v}$ from $\rho_k^t(\mathbf{x}|\mathbf{v})$, whereas at lower levels, $\mathbf{v}_k$ is a *differential* velocity related with the likelihood estimation $\tilde{\rho}_k^t(\mathbf{x}|\mathbf{v}_k)$. Note also that ρ_k^t correlates $\tilde{\mathbf{G}}_k^{t,\mathbf{x}}$ (and not $\mathbf{G}_k^{t,\mathbf{x}}$) with $\mathbf{G}_k^{t+\Delta t,\mathbf{x}+\Delta\mathbf{x}}$.
3. Calculate the local likelihood for the *physical* velocity $\mathbf{v}$ by combining the estimation for the physical velocity from the higher stage $k-1$ with the likelihood estimations from stage k,
$$\rho_k^t(\mathbf{x}|\mathbf{v} = \mathbf{v}_{k-1} + \mathbf{v}_k) :\sim \sum_{\mathbf{v}_k,\mathbf{v}_{k-1}} \tilde{\rho}_k^t(\mathbf{x} + \mathbf{v}_{k-1}\Delta t|\mathbf{v}_k)\, \rho_{k-1}^t(\mathbf{v}_{k-1}|\mathbf{x}) \ . \tag{11}$$
At the highest level there will be no combination because no velocity distributions from a coarser level are available and therefore $\rho_0^t(\mathbf{x}|\mathbf{v}) := \tilde{\rho}_0^t(\mathbf{x}|\mathbf{v})$.
4. Combine the likelihood with the prior $\rho_k(\mathbf{v}_k)$ to get the local a-posteriory probability for the *physical* velocity $\mathbf{v}$ according to
$$\rho_k^t(\mathbf{v}|\mathbf{x}) \sim \rho_k^t(\mathbf{x}|\mathbf{v})\, \rho_k(\mathbf{v}) \ . \tag{12}$$
5. Use the gained a-posteriory probability for the prediction of the image at time $t + \Delta t$ at the next level $k + 1$ according to
$$\tilde{\mathbf{G}}_{k+1}^t := \sum_{\mathbf{v},\mathbf{x}} \rho_k^t(\mathbf{v}|\mathbf{x}) \mathbf{W}^{\mathbf{x}-\mathbf{v}\Delta t} \odot \mathbf{G}_{k+1}^t \ . \tag{13}$$
This is the best estimate according to level k and the constraints given by the generative model. $\mathbf{W}^{\mathbf{x}-\mathbf{v}\Delta t}$ is the window shifted by $\mathbf{x} - \mathbf{v}\Delta t$ and takes into account the correct window weightings.
6. Increase the pyramid level k and return to point 1.

4 Results

The results of the hierarchical procedure in Fig. 3 show that a combination of velocity distributions is possible within a velocity resolution pyramid and that the process combines advantages of the different levels of resolution. The coarser

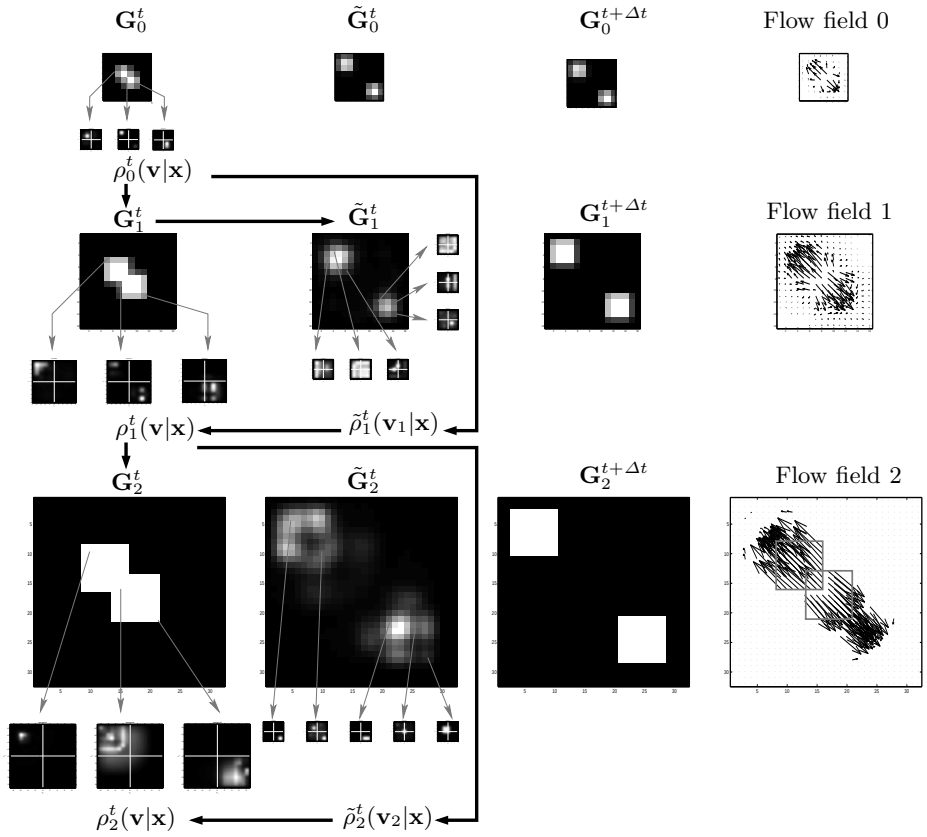

Fig. 2. The results of the hierarchical velocity distribution calculations using the resolution pyramid. The right column shows flow fields (extracted for evaluation purposes). The left three columns show the original images at time t (first column) and $t + \Delta t$ (third column), as well as the reconstructed image $\tilde{\mathbf{G}}_k^t$ (second column), which is the *prediction* of image at time $t + \Delta t$ (third column) using the available velocity distributions at time t.At each resolution level, a number of representative 2D velocity distributions $\rho_k^t(\mathbf{v}|\mathbf{x})$ (absolute velocities $\mathbf{v}$) and $\rho_k^t(\mathbf{v}_k|\mathbf{x})$ (relative velocities $\mathbf{v}_k$) are shown, with gray arrows indicating their positions in the image. For the distributions, white lines indicate the velocity coordinate system, with black/white representing low/high probabilities for the corresponding velocity. Black arrows indicate the order of the computations within the pyramid. The coarse-to-fine calculation allows the system to use the coarse velocity information for regions which at higher levels of resolution have flat distributions or ambiguous motion signals, and to refine the coarse information with additional information from the finer levels of velocity resolution. This can be seen at the flow field at the highest level of resolution ("Flow field 2"). In "Flow field 2" only the velocity vectors with high probabilities are shown.

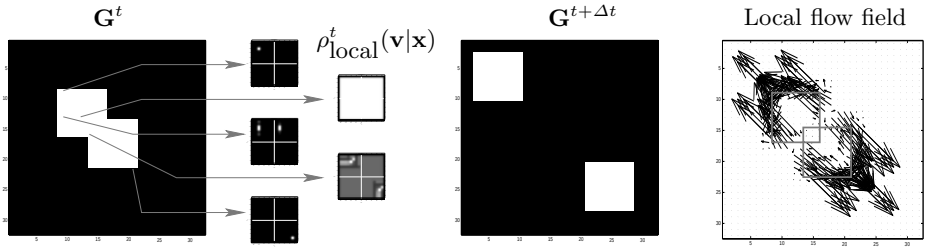

Fig. 3. For comparison, here we show the velocity distributions and the flow field calculated over the entire velocity range using fine velocity discretization and fine resolution window. In the flow field we see that the velocity signals are only unambiguous at the corners of the squares, whereas the central parts convey no motion information and the edges suffer the classical aperture problem. Using the squared magnitude of the difference between the correct and estimated flow the pyramidal approach produces 54,4% less error than this one.

levels of the pyramid analyze larger patches and provide estimations for larger velocities. Nevertheless, the estimations are often inaccurate, dependent on the shape of the velocity distributions. In contrast, the finer levels of the pyramid operate more locally and analyze smaller velocities. This leads in some cases to peaked velocity distributions, but in other cases (e.g., when there is not sufficient structure) to broad distributions because of unavailable motion information. The combination of coarse and fine levels using the velocity distribution representation allows to incorporate more global velocity estimations if local information is missing, and to refine global velocity estimations if local information is present. An advantage of a pyramidal structure for velocity computation is that we gain the coarse estimations very fast, and can then refine the results step by step. The strategy is comparable to detecting global motions first, and then to use this information in a moving coordinate frame, in order to detect the finer relative motions still available within this frame.

References

1. *Handbook of Computer Vision and Applications*, chapter Bayesian Multi-Scale Differential Optical Flow. Academic Press, 1999.
2. J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation.
3. E. Simoncelli, E.H. Adelson, and D.J. Heeger. Probability distributions of optical flow. In *Proc Conf on Computer Vision and Pattern Recognition*, pages 310–315, Mauii, Hawaii, 1991. IEEE Computer Society.
4. Y. Weiss and D.J. Fleet. Velocity likelihoods in biological and machine vision.
5. Qing X. Wu. A correlation-relaxation-labeling framework for computing optical flow - template matching from a new perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995.