Optical Flow Estimation from Monogenic Phase

Michael Felsberg*

Linköping University, Computer Vision Laboratory,
SE-58183 Linköping, Sweden,
mfe@isy.liu.se, <http://www.isy.liu.se/cvl/>

Abstract. The optical flow can be estimated by several different methods, some of them require multiple frames some make use of just two frames. One approach to the latter problem is optical flow from phase. However, in contrast to (horizontal) disparity from phase, this method suffers from the phase being oriented, i.e., classical quadrature filter have a predefined orientation in which the phase estimation is correct and the phase error grows with increasing deviation from the local image orientation. Using the approach of the monogenic phase instead, results in correct phase estimates for all orientations if the signal is locally 1D. This allows to estimate the optical flow with sub-pixel accuracy from a multi-resolution analysis with seven filter responses at each scale. The paper gives a short and easy to comprehend overview about the theory of the monogenic phase and the formula for the displacement estimation is derived from a series expansion of the phase. Some basic experiments are presented.

1 Introduction

The aim of this paper is not only to present just another optical flow estimation approach, but also to give a tutorial-like introduction to the topic of the monogenic signal and its phase. We try to strip off all theoretic background which is unnecessary and focus on a simple, concrete, and complete description of the framework. For a more formal treatment, we refer to the earlier publications on the monogenic framework [1,2,3].

Based on the monogenic phase, we then derive a simple formula for measuring oriented displacements between two images, which is then applied in a multi-scale approach for the estimation of the optical flow. This method is quite similar to the one presented in [4], with two differences: We do not know the direction of displacement in advance and we show the pure estimates, i.e., we do not post-process the point estimates with channel smoothing. Combining the point-wise estimator and a non-linear smoothing technique easily compensates errors in the estimates and partly the aperture effect, but it was not our aim to present a complete motion estimation system but rather the signal processing part of it.

* This work has been supported by EC Grant IST-2003-004176 COSPAL and by EC Grant IST-2002-002013 MATRIS.

2 Why is Phase-Based Image Processing Preferable?

Before we start to introduce the framework of the monogenic phase, we motivate why we want to use phase-based methods at all. Basically, there are three reasons:

1. There is strong evidence that the human visual system makes use of local phase in V1 [5].
2. Phase-based processing is to a large extent invariant to changes of lighting conditions. The local image intensity can additionally be used to measure the reliability of measurements.
3. Perceptually, the reconstruction of an image from phase information is much better than that from amplitude information.

Reasons 2 and 3 have to be explained in some more detail. Note that we always consider a local region in the spatial-frequency domain, i.e., we look at local image regions at certain frequency ranges (or equivalently: at certain scales). Note in this context that we denote spatial frequency or wavenumber by 'frequency', not a frequency in the temporal sense.

2.1 The Image Model of Local Phase

Our image model that we apply in phase-based processing is

$$I(\mathbf{x}) = \tilde{A}(\mathbf{x}) \cos(\tilde{\varphi}(\mathbf{x})) + \bar{I} \quad (1)$$

where $\mathbf{x} = (x_1, x_2)^T$ indicates the spatial coordinate vector, $I(\mathbf{x})$ the image, $\bar{I}$ the average intensity (DC level), $\tilde{A}(\mathbf{x})$ the instantaneous amplitude (real-valued, non-negative), and $\tilde{\varphi}(\mathbf{x})$ the instantaneous phase [6]. The average intensity is irrelevant for the analysis of the image contents and in the human visual system it is largely compensated already during the image acquisition. In this model the decomposition into $\tilde{A}(\mathbf{x})$ and $\tilde{\varphi}(\mathbf{x})$ seems to be highly ambiguous. This is however not the case, since the amplitude is a non-negative real number. Hence, the zeros of $I(\mathbf{x}) - \bar{I}$ must be covered by zeros of $\cos(\tilde{\varphi}(\mathbf{x}))$. Assuming sufficiently smooth functions, the zero crossings are in direct correspondence to the full phase [7] and the instantaneous phase becomes a uniquely defined feature.

If we switch to a local region in the spatial-frequency domain, i.e., we consider an image region and a small range of frequencies, the model (1) becomes much simpler. Under the assumption of small magnitude variations of the considered frequency components in the local spectrum, the amplitude becomes approximately constant in the local region. It is therefore referred to as the local amplitude $A_{\mathbf{x}}$, where $\mathbf{x}$ now indicates the origin of the local region, i.e., all estimates with subscript $\mathbf{x}$ refer to the region with origin $\mathbf{x}$. The model (1) becomes for the local region

$$\tilde{I}(\mathbf{x} + \mathbf{y}) = A_{\mathbf{x}} \cos(\varphi_{\mathbf{x}}(\mathbf{y})) \quad , \quad (2)$$

where $\tilde{I}(\mathbf{x} + \mathbf{y})$ is the local image patch, $\mathbf{y}$ the local patch coordinate vector, and $\varphi_{\mathbf{x}}(\mathbf{y})$ is the local phase.

Assuming a small range of frequencies, the local phase cannot vary arbitrarily fast, it has a high degree of smoothness. Therefore, it can be well approximated by a first order series in $\mathbf{y} = 0$:

$$\varphi_{\mathbf{x}}(\mathbf{y}) \approx \mathbf{f}_{\mathbf{x}}^T \mathbf{y} + \varphi_{\mathbf{x}}(0) = f_{\mathbf{x}} \mathbf{n}_{\mathbf{x}}^T \mathbf{y} + \varphi_{\mathbf{x}}(0) , \quad (3)$$

where $f_{\mathbf{x}}$ is the local frequency, $\mathbf{n}_{\mathbf{x}}$ is the local orientation (unit vector), and $\varphi_{\mathbf{x}}(0)$ is some phase offset. That means, our local image model directly led to the assumption of intrinsic dimensionality one [8], or simple signals [6], where $\tilde{I}(\mathbf{x} + \mathbf{y}) = \hat{I}_{\mathbf{x}}(\mathbf{n}_{\mathbf{x}}^T \mathbf{y})$ ($\hat{I}$ being a suitable 1D function).

We can group the series of the local phase (3) in two different ways:

$$\varphi_{\mathbf{x}}(\mathbf{y}) \approx f_{\mathbf{x}}(\mathbf{n}_{\mathbf{x}}^T \mathbf{y}) + \varphi_{\mathbf{x}}(0) = \bar{\varphi}_{\mathbf{x}}(\mathbf{n}_{\mathbf{x}}^T \mathbf{y}) \quad \text{and} \quad (4)$$

$$\varphi_{\mathbf{x}}(\mathbf{y}) \approx \mathbf{n}_{\mathbf{x}}^T (\mathbf{n}_{\mathbf{x}} \mathbf{n}_{\mathbf{x}}^T \mathbf{y} f_{\mathbf{x}} + \mathbf{n}_{\mathbf{x}} \varphi_{\mathbf{x}}(0)) = \mathbf{n}_{\mathbf{x}}^T \mathbf{r}_{\mathbf{x}}(\mathbf{y}) . \quad (5)$$

Whereas the former expression is a 1D function with a scalar product as an argument, the latter expression is a 2D vector field, the local phase vector $\mathbf{r}_{\mathbf{x}}$, which is projected onto the orientation vector. Although the distinction seems to be trivial, the local phase vector is simpler to estimate, because we do not need to know the local orientation in advance. We will return to this issue later.

2.2 Lighting Invariance and Perceptual Image Contents

Before we present an estimation procedure for the local phase, we continue the discussion on the reasons 2 and 3 for using phase-based methods. The decomposition of an image into local amplitude and local phase at a particular frequency range means to neglect the high frequencies, to represent the intermediate frequencies by the local phase, and to cover the lower frequencies by (more global) changes of the local amplitude. Changes of lighting conditions are then to a large extend represented by changes of the local amplitude, with the exception of moving shadow boundaries. Hence, most changes of lighting conditions are not visible in the phase representation, cf. Fig. 1.

Since the famous paper [9], it is well known that the reconstruction from the phase spectrum is much better from a perceptual point of view than the one from the magnitude spectrum. Considering the Fourier spectrum means to consider the part of the spatial-frequency domain with maximum localization in frequency and no localization in position. If we move to some point with finite localization in both spaces, the results of the experiment from [9] still remain valid, cf. Fig. 2, although we now consider the local phase.

If the image is decomposed into its amplitude and phase information, it becomes evident that the local amplitude is basically just a measure for the confidence of the extracted phase, i.e., in technical terms it represents the signal-to-noise ratio (SNR), cf. Fig. 2, center column. The local phase represents most of the image structure, cf. Fig. 2, left column. In the areas where the amplitude is close to zero, thus meaning 'no confidence', the local phase contains mainly noise. In the regions of non-zero confidence, the cosine of the local phase results in a visual impression which comes very close to the bandpass filtered image, cf. Fig. 2, right column.

Fig. 1. The checker shadow illusion 'disillusioned'. Left: original image from http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html. Right: reconstruction from local phase.

Fig. 2. Decomposing an image into its local phase and its local amplitude. From left to right: $\cos(\varphi_{\mathbf{x}_0})$, $A_{\mathbf{x}_0}$, $\tilde{I}(\mathbf{x}_0)$, where the intensities are adjusted to obtain similar intensity ranges. Grey means zero, white means positive values, and black means negative values. Top row: full size images. Bottom row: image detail. $\tilde{I}(\mathbf{x}_0)$ is obtained from $I(\mathbf{x}_0)$ applying the filters from [10] using scales $\{3,4,5\}$.

3 The Monogenic Signal: A Survey

The monogenic signal provides a framework to estimate the local phase, the local orientation, and the local amplitude of an image [3]. It can be considered as a 2D generalization of the analytic signal. Former 2D generalizations tried to estimate the local phase according to (4), which was only partly successful since the local orientation has to be known in advance to steer the filters [11,12]. In the monogenic framework, however, we try to estimate the local phase vector (5) instead, where the local orientation is part of the result and need not be known in advance. The estimated vector has a natural upper bound for its error [1].

3.1 Spherical Quadrature Filter

Quadrature filters in 1D are constructed in two steps:

1. Select a suitable bandpass filter which is responsible for the localization in the time-frequency domain, i.e., it responds to signal contributions which are in a certain frequency range (the passband) and in a certain time window. This bandpass filter is an even filter, i.e., it is symmetric.
2. Compute the Hilbert transform of the bandpass filter in order to construct the corresponding odd, i.e., antisymmetric, filter which has the same magnitude spectrum.

Practical problems concerning computing the Hilbert transform are out of the scope of this paper. The quadrature filter pair is mostly applied as a complex filter to the 1D signal. The response is a complex signal, which is divided into magnitude and argument. One can easily show that the argument is an estimate for the local phase. In Fig. 3 left, the 1D phase interpretation is illustrated. The figure is generated by projecting all possible phase-responses onto the filter. As a result we get those input signals which would generate the in-fed responses. Keeping the amplitude constant and varying the phase from 0 to 2π results in the sketched signal profiles (at 1, i , -1 , and $-i$) and the continuously varying intensities in the background. With increasing phase angle, the quadrature filter turns from a purely even filter towards a purely odd filter. Continuing further than $\pi/2$ leads towards an even filter again, but with opposite sign. After π the filter becomes first odd (opposite sign) and finally it turns back into the initial filter. The corresponding signal profile changes from a positive impulse over a positive step (from inside to outside), over a negative impulse, and over a negative step, until it is a positive impulse again.

The 2D spherical quadrature filters (SQF) are constructed likewise:

1. Select a suitable *radial* bandpass filter, i.e., a rotation invariant filter. The passband consists of frequencies of a certain range in absolute value, but with arbitrary direction. This bandpass filter is an even filter, i.e., it is symmetric.
2. Compute the *Riesz* transform of the bandpass filter in order to construct the corresponding odd, i.e., antisymmetric about the origin, filter which has the same magnitude spectrum. The odd filter consists of two components.

Fig. 3. Local phase models. Left: the 1D phase, the corresponding filter shapes at 1, i , -1 , and $-i$, and the continuously changing signal profile (grey values in the background). Right: the local phase of the monogenic signal. The 3D vector $(q_1, q_2, p)^T$ together with the p -axis define a plane at orientation θ in which the rotation takes place. The normal of this plane multiplied by the rotation angle φ results in the rotation vector $\mathbf{r}^\perp$.

The radial bandpass filter is given by some suitable frequency response $B_e(\rho)$, where (ρ, ϕ) are the polar coordinates of the frequency domain, such that it is rotational symmetric and therefore symmetric (even) about the origin. The corresponding antisymmetric (odd) filters are then given by

$$B_{o1}(\rho, \phi) = i \cos \phi B_e(\rho) \quad \text{and} \quad B_{o2}(\rho, \phi) = i \sin \phi B_e(\rho) . \quad (6)$$

All together, an SQF provides three responses; the even filter response $p(\mathbf{x}) = (I * b_e)(\mathbf{x})$ and the two odd filter responses $\mathbf{q}(\mathbf{x}) = (q_1(\mathbf{x}), q_2(\mathbf{x}))^T = ((I * b_{o1})(\mathbf{x}), (I * b_{o2})(\mathbf{x}))^T$.

3.2 Extracting Local Phase

The local amplitude can be extracted likewise as in the 1D case by calculating the magnitude of the 3D vector:

$$A_{\mathbf{x}} = \sqrt{q_1(\mathbf{x})^2 + q_2(\mathbf{x})^2 + p(\mathbf{x})^2} , \quad (7)$$

cf. Fig. 4 for an example. The phase, however, cannot be extracted as the argument of a complex number, since we need two angles to describe the 3D rotation from a reference point (on the p -axis) into the SQF response. These angles are indicated in Fig. 3, and they have direct interpretations in terms of local orientation and local phase.

Fig. 4. Upper row from left to right: original signal, local amplitude (in both cases white means zero and black means large values), local phase according to (4) (black means negative phase, white means positive phase). Bottom row left: local orientation. Right: local phase vector. The length of the phase vector is encoded in grey values. The local orientation and phase are only displayed for sufficiently large amplitude.

It has been shown in [3] that an image patch with intrinsic dimensionality one and local orientation θ (w.r.t. the horizontal axis) results in a response of the form $(q_1(\mathbf{x}), q_2(\mathbf{x}), p(\mathbf{x}))^T = (\cos \theta q(\mathbf{x}), \sin \theta q(\mathbf{x}), p(\mathbf{x}))^T$ for a suitable $q(\mathbf{x})$, i.e., according to Fig. 3, the rotation takes place in a plane which encloses angle θ with the (q_1, p) -plane. For non-zero $\mathbf{q}$ this angle can hence be estimated as

$$\theta_{\mathbf{x}} = \tan^{-1} \left(\frac{q_2(\mathbf{x})}{q_1(\mathbf{x})} \right) \in (-\pi/2; \pi/2] , \quad (8)$$

where an orientation – direction ambiguity occurs, since the directional sense of a 2D signal cannot be extracted from a local signal [6], i.e., $\mathbf{q}$ and $-\mathbf{q}$ map onto the same orientation, cf. Fig. 4.

A further result from [3] is the connection between Hilbert transform and Riesz transform, which practically means that a 1D projection of an SQF results in a 1D quadrature filter. Since a signal with intrinsic dimensionality of one is constant along a certain orientation, it leads to a 1D projection of the involved filter kernels. Hence, p and q from the previous paragraph can be considered as a 1D quadrature filter response and the local phase is given by $\arg(p + iq)$.

However, we do not know the correct sign of q , since it depends on the directional sense of θ . The best possible solution is to project $\mathbf{q}$ onto $(\cos \theta, \sin \theta)$:

$$\bar{\varphi} = \arg(p + i(\cos \theta q_1 + \sin \theta q_2)) = \arg(p + i \text{sign}(q_1) |\mathbf{q}|) , \quad (9)$$

see also Fig. 4. The sign depends on q_1 , because $\theta \in (-\pi/2; \pi/2]$ which corresponds to the quadrants with positive q_1 , i.e., $(\cos \theta, \sin \theta) = \text{sign}(q_1) \mathbf{q}/|\mathbf{q}|$. The derived phase estimate is actually an estimate according to (4), since $\mathbf{q}$ can be considered as a steerable filter projected onto $\mathbf{n} = (\cos \theta, \sin \theta)$.

In order to obtain a continuous representation of orientation and phase, both are combined in the phase vector

$$\mathbf{r} = \bar{\varphi} (\cos \theta, \sin \theta)^T = \frac{\mathbf{q}}{|\mathbf{q}|} \arg(p + i|\mathbf{q}|) , \quad (10)$$

cf. Fig. 3, right, and Fig. 4. Note that the rotation vector $\mathbf{r}^\perp$ is perpendicular to the local phase vector and the local orientation. The phase vector $\mathbf{r}$ is an estimate according to (5) and we will use it subsequently instead of the scalar phase $\bar{\varphi}$.

3.3 Estimating the Local Frequency

As pointed out above, the local phase model was derived from a first order series expansion (3) which contained the local frequency $f_{\mathbf{x}}$. The latter can also be estimated directly from the signal, using the spatial derivatives of (10). Since the first order approximation leads to an intrinsic dimensionality of one, we only need to consider the directional derivative along $\mathbf{n}_{\mathbf{x}}$ w.r.t. $\mathbf{y}$ ($\nabla = (\partial_{y_1}, \partial_{y_2})$):

$$\bar{f}_{\mathbf{x}} = (\mathbf{n}_{\mathbf{x}}^T \nabla) \varphi_{\mathbf{x}}(\mathbf{y}) \approx (\mathbf{n}_{\mathbf{x}}^T \nabla)(\mathbf{n}_{\mathbf{x}}^T \mathbf{r}_{\mathbf{x}}(\mathbf{y})) = \nabla^T \mathbf{n}_{\mathbf{x}} \mathbf{n}_{\mathbf{x}}^T \mathbf{r}_{\mathbf{x}}(\mathbf{y}) = \nabla^T \mathbf{r}_{\mathbf{x}}(\mathbf{y}) , \quad (11)$$

where the last step is correct, since $\mathbf{r}_{\mathbf{x}}(\mathbf{y}) = \mathbf{n}_{\mathbf{x}} \mathbf{n}_{\mathbf{x}}^T \mathbf{y} f_{\mathbf{x}} + \mathbf{n}_{\mathbf{x}} \varphi_{\mathbf{x}}(0)$, i.e., collinear with $\mathbf{n}_{\mathbf{x}}$, cf. (5).

Hence, the divergence of the local phase vector yields the local frequency. However, we do not even need to compute these derivatives explicitly, which would by the way result in some trouble with the wraparounds [13]. Instead, we do some calculus on (10) (where we leave out some indices for convenience):

$$\begin{aligned} \nabla^T \mathbf{r} &= \nabla^T \frac{\mathbf{q}}{|\mathbf{q}|} \arg(p + i|\mathbf{q}|) \\ &\stackrel{\mathbf{q}/|\mathbf{q}| = \pm \mathbf{n}}{=} \frac{\mathbf{q}^T}{|\mathbf{q}|} \nabla \arg(p + i|\mathbf{q}|) \\ &\stackrel{\partial \arg = \partial \tan^{-1}}{=} \frac{\mathbf{q}^T}{|\mathbf{q}|} \nabla \tan^{-1} \left(\frac{|\mathbf{q}|}{p} \right) \\ &= \frac{\mathbf{q}^T}{|\mathbf{q}|} \frac{1}{1 + \frac{|\mathbf{q}|^2}{p^2}} \frac{p \nabla |\mathbf{q}| - |\mathbf{q}| \nabla p}{p^2} \\ &= \frac{\mathbf{q}^T}{|\mathbf{q}|} \frac{1}{p^2 + |\mathbf{q}|^2} \frac{p(\nabla \mathbf{q}^T) \mathbf{q} - |\mathbf{q}|^2 \nabla p}{|\mathbf{q}|} . \end{aligned} \quad (12)$$

Since we know that we deal with an intrinsically 1D region, $\mathbf{q}$ itself is also 1D, and we can write it as

$$\mathbf{q}(\mathbf{y}) = \mathbf{n}q(\mathbf{n}^T \mathbf{y}) ,$$

such that

$$\mathbf{q}^T (\nabla \mathbf{q}^T) \mathbf{q} = |q|^2 \mathbf{n}^T (\nabla \mathbf{n}^T q) \mathbf{n} = |q|^2 \mathbf{n}^T \nabla q = |\mathbf{q}|^2 \nabla^T \mathbf{q} ,$$

and finally

$$\nabla^T \mathbf{r} = \frac{p \nabla^T \mathbf{q} - \mathbf{q}^T \nabla p}{p^2 + |\mathbf{q}|^2} . \quad (13)$$

By this expression we can directly estimate the local frequency $\bar{f}_{\mathbf{x}}$ by a quotient consisting of the three filter responses and their four partial derivatives $\partial_1 p$, $\partial_2 p$, $\partial_1 q_1$, and $\partial_2 q_2$.

3.4 A Concrete Filter Set

Up to now, we have not specified a particular set of SQF. Suitable filter sets have to be chosen according to the application. In most cases, the radial bandpass filter B_e will be designed in the Fourier domain. The frequency responses for the corresponding two other filters are given by means of (6). The spatial derivatives of these filters are obtained by multiplying the frequency responses of the former with $i\rho \cos \phi$ resp. $i\rho \sin \phi$ according to the derivative theorem of the Fourier transform [6]. The impulse responses of all filters are computed by numerical optimization, see e.g. [6].

This optimization becomes unnecessary if the inverse Fourier transforms of all frequency responses are known. This is the case for filters based on the Poisson filter series [10] which have been used for the following experiments. Starting point are the Poisson filters and their Riesz transforms (lower case: spatial domain, upper case: frequency domain, Fourier transform according to [6]):

$$h_e(\mathbf{x}; s) = \frac{s}{2\pi(s^2 + |\mathbf{x}|^2)^{3/2}} \quad H_e(\rho, \phi; s) = \exp(-\rho s) \quad (14)$$

$$h_{o1}(\mathbf{x}; s) = \frac{x_1}{2\pi(s^2 + |\mathbf{x}|^2)^{3/2}} \quad H_{o1}(\rho, \phi; s) = i \cos \phi \exp(-\rho s) \quad (15)$$

$$h_{o2}(\mathbf{x}; s) = \frac{x_2}{2\pi(s^2 + |\mathbf{x}|^2)^{3/2}} \quad H_{o2}(\rho, \phi; s) = i \sin \phi \exp(-\rho s) . \quad (16)$$

The SQF for the Poisson filter series with zero value and zero first derivative at $\rho = 0$ is given for arbitrary scale $s > 1$ by

$$b_e(\mathbf{x}; s) = h_e(\mathbf{x}; s-1) - 2h_e(\mathbf{x}; s) + h_e(\mathbf{x}; s+1) \quad (17)$$

$$b_{o1}(\mathbf{x}; s) = h_{o1}(\mathbf{x}; s-1) - 2h_{o1}(\mathbf{x}; s) + h_{o1}(\mathbf{x}; s+1) \quad (18)$$

$$b_{o2}(\mathbf{x}; s) = h_{o2}(\mathbf{x}; s-1) - 2h_{o2}(\mathbf{x}; s) + h_{o2}(\mathbf{x}; s+1) . \quad (19)$$

The derivatives of these filters are straightforward to compute in either domain.

4 Optical Flow Estimation

In this section we will propose a method for two-frame optical flow estimation. Optical flow estimation is the first step towards motion estimation. The optical flow might differ essentially from the motion field, but this aspect is out of the scope of this paper. There are several methods for optical flow estimation known from the literature, among these a phase-based method for the two-frame case, see e.g. [13]. We will adapt this approach for the monogenic framework.

4.1 Local Displacements

The main idea of flow from phase is to make a local displacement estimation by a first order series expansion of the phase, which was a popular disparity estimation technique in the 90ies [14,15,16]. This series has to be slightly adapted for the phase vector [4]. We start with the assumption that the new frame is obtained from the old one by a local displacement $\mathbf{d}(\mathbf{x})$:

$$I_{\text{new}}(\mathbf{x}) = I_{\text{old}}(\mathbf{x} - \mathbf{d}(\mathbf{x})) . \quad (20)$$

In our local image model, this assumption maps directly to the phase vectors:

$$\mathbf{r}_{\text{new}}(\mathbf{x}) = \mathbf{r}_{\text{old}}(\mathbf{x} - \mathbf{d}(\mathbf{x})) . \quad (21)$$

Assuming $|\mathbf{d}(\mathbf{x})|$ is sufficiently small, we can approximate the phase of the new frame by a first order expansion (c.f. (5)):

$$\mathbf{r}_{\text{new}}(\mathbf{x}) \approx \mathbf{r}_{\text{old}}(\mathbf{x}) - \mathbf{n}(\mathbf{x})\mathbf{n}^T(\mathbf{x})\mathbf{d}(\mathbf{x})f(\mathbf{x}) . \quad (22)$$

Rearranging this equation and assuming a constant displacement in the neighborhood $\mathcal{N}$, we obtain $\mathbf{d}_{\mathcal{N}}$ from the linear system

$$\sum_{\mathbf{x} \in \mathcal{N}} \mathbf{r}_{\text{diff}}(\mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{N}} [\mathbf{r}_{\text{old}}(\mathbf{x}) - \mathbf{r}_{\text{new}}(\mathbf{x})] = \sum_{\mathbf{x} \in \mathcal{N}} [\mathbf{n}(\mathbf{x})\mathbf{n}^T(\mathbf{x})f(\mathbf{x})]\mathbf{d}_{\mathcal{N}} , \quad (23)$$

where the phase vector difference is modulo π (concerning the vector magnitude). In practice, the phase vector difference is calculated from the 3D rotation which relates the two SQF responses. This can be done easiest in the algebra of quaternions, resulting in the following four components:

$$\begin{aligned} p_{\text{diff}} &= p_{\text{old}}p_{\text{new}} + \mathbf{q}_{\text{old}}^T \mathbf{q}_{\text{new}} \\ \mathbf{q}_{\text{diff}} &= p_{\text{old}}\mathbf{q}_{\text{new}} - \mathbf{q}_{\text{old}}p_{\text{new}} \\ c_{\text{diff}} &= \mathbf{q}_{\text{old}}^T \mathbf{q}_{\text{new}}^\perp . \end{aligned}$$

The first three components describe the local displacement, i.e., extracting the 'phase vector' from $(\mathbf{q}_{\text{diff}}^T, p_{\text{diff}})^T$ using (10) yields the phase difference for (23). Note that the amplitude of $(\mathbf{q}_{\text{diff}}^T, p_{\text{diff}})^T$ can be used as a measurement of reliability.

The fourth component is somehow special: it describes the rotation of the local orientation. This rotation angle is obtained as

$$\psi(\mathbf{x}) = \sin^{-1} \left(\frac{c_{\text{diff}}}{|\mathbf{q}_{\text{new}}||\mathbf{q}_{\text{old}}|} \right) . \quad (24)$$

The two fields, i.e., the optical flow $\mathbf{i}(\mathbf{x})$ and the local rotation field $\psi(\mathbf{x})$ are not independent, but they can be used in parallel to enhance the consistency of the estimated fields, cf. also [17].

Note that (23) is directly related to the brightness constancy constraint equation (BCCE). If the intensity in the BCCE is replaced with the phase vector, i.e., constraining phase vector constancy, and by approximating the time derivative with a finite difference, we obtain (23).

4.2 Some Examples

There is definitely a need for detailed experiments and comparisons to other methods, but at the current state, we only present some basic experiments on synthetic data with known ground truth, see Fig. 5, and the taxi sequence, see Fig. 6. In the synthetic experiments, we have applied constant translational motion to the synthetic pattern.

For the taxi sequence, cf. Fig. 6, we have applied a single scale estimation. This was possible, since the flow field has small magnitudes (< 3), such that the convergence radius of the applied filter was sufficient. Note that no post-processing has been applied to the estimates, i.e., the only regularization is given by the applied bandpass filter (here: $s = 2$).

Fig. 5. Flow estimation experiment with synthetic pattern. Left: pattern (Siemens star). Right: absolute error for constant motion estimate with windowed averaging (binomial filter) in (23). The true motion vector was $(\sqrt{2}, \sqrt{3})$.

Fig. 6. Experiments on the taxi sequence. Upper left: one frame of the sequence. Upper right: confidences of the estimates in logarithmic scale (white means high confidence). Bottom: estimated flow field for a smaller region of the frame.

5 Summary and Outlook

We have presented a self-contained survey on the monogenic framework and derived optical flow estimation as an application of it. The method needs, however, further investigation concerning the quality of the estimates in comparison to other two-frame methods. For obtaining a full motion estimation system, the method has to be combined with some appropriate post-processing in order to replace the linear averaging in (23), see e.g. [18].

References

1. Felsberg, M., Duits, R., Florack, L.: The monogenic scale space on a rectangular domain and its features. *International Journal of Computer Vision* (2004) accepted for publication in a future issue.
2. Felsberg, M., Sommer, G.: The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. *Journal of Mathematical Imaging and Vision* **21** (2004) 5–26
3. Felsberg, M., Sommer, G.: The monogenic signal. *IEEE Transactions on Signal Processing* **49** (2001) 3136–3144
4. Felsberg, M.: Disparity from monogenic phase. In v. Gool, L., ed.: 24. DAGM Symposium Mustererkennung, Zürich. Volume 2449 of *Lecture Notes in Computer Science*, Springer, Heidelberg (2002) 248–256
5. Mechler, F., Reich, D.S., Victor, J.D.: Detection and discrimination of relative spatial phase by V1 neurons. *Journal of Neuroscience* **22** (2002) 6129–6157
6. Granlund, G.H., Knutsson, H.: *Signal Processing for Computer Vision*. Kluwer Academic Publishers, Dordrecht (1995)
7. Hurt, N.E.: *Phase Retrieval and Zero Crossings*. Kluwer Academic, Dordrecht (1989)
8. Krieger, G., Zetsche, C.: Nonlinear image operators for the evaluation of local intrinsic dimensionality. *IEEE Transactions on Image Processing* **5** (1996) 1026–1041
9. Oppenheim, A., Lim, J.: The importance of phase in signals. *Proc. of the IEEE* **69** (1981) 529–541
10. Felsberg, M.: On the design of two-dimensional polar separable filters. In: 12th European Signal Processing Conference, Vienna, Austria. (2004)
11. Knutsson, H., Wilson, R., Granlund, G.H.: Anisotropic non-stationary image estimation and its applications: Part I – restoration of noisy images. *IEEE Trans. on Communications* **COM-31** (1983) 388–397
12. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13** (1991) 891–906
13. Jähne, B.: *Digital Image Processing*. Springer, Berlin (2002)
14. Fleet, D.J., Jepson, A.D., Jenkin, M.R.M.: Phase-based disparity measurement. *Computer Vision, Graphics, and Image Processing. Image Understanding* **53** (1991) 198–210
15. Maimone, M.W., Shafer, S.A.: Modeling foreshortening in stereo vision using local spatial frequency. In: *International Robotics and Systems Conference*, IEEE Computer Society Press (1995) 519–524
16. Maki, A., Uhlin, T., Eklundh, J.O.: A direct disparity estimation technique for depth segmentation. In: *Proc. 5th IAPR Workshop on Machine Vision Applications*. (1996) 530–533
17. Schnörr, C.: Variational methods for fluid flow estimation. In Jähne, B., Barth, E., Mester, R., Scharr, H., eds.: *Complex Motion, Proceedings 1st Int. Workshop*, Günzburg, Springer Verlag, Heidelberg (2004)
18. Forssén, P.E.: Multiple motion estimation using channel matrices. In Jähne, B., Barth, E., Mester, R., Scharr, H., eds.: *Complex Motion, Proceedings 1st Int. Workshop*, Günzburg (2004)