

# Disparity from Monogenic Phase\*

Michael Felsberg

Department of Electrical Engineering, Linköping University,  
SE-58183 Linköping, Sweden

`mfe@isy.liu.se`,

WWW home page: <http://www.isy.liu.se/~mfe>

**Abstract.** Disparity estimation is a fundamental problem of computer vision. Besides other approaches, disparity estimation from phase information is a quite wide-spread technique. In the present paper, we have considered the influence of the involved quadrature filters and we have replaced them with filters based on the monogenic signal. The implemented algorithm makes use of a scale-pyramid and applies channel encoding for the representation and fusion of the estimated data. The performed experiments show a significant improvement of the results.

## 1 Disparity Estimation

In this paper we introduce a new method for estimating the disparity between two images. Disparity typically occurs in a stereo camera setting. Due to the depth in space, the image plane position of an object point differs in the left and the right image. Knowing this displacement, it is possible to determine the depth of the object point and therefore its 3D position if the camera has been calibrated beforehand. Neglecting the problems of calibration and back-projection into 3D space, we solely focus on the problem of estimating the disparity.

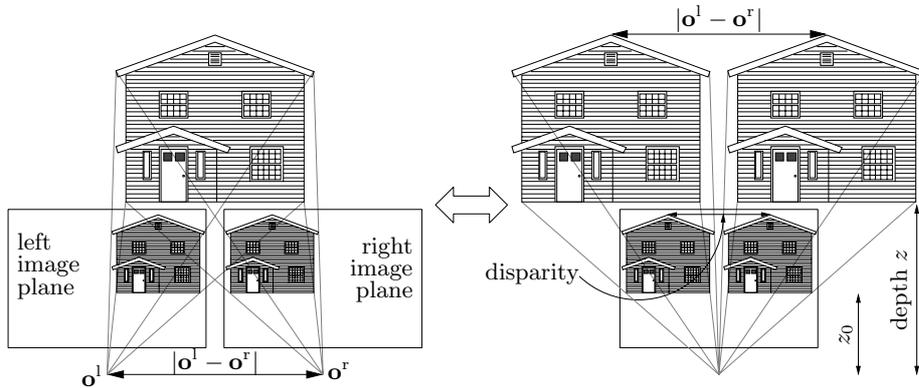
### 1.1 Geometric Settings

The geometric setting of a stereo system is sketched in Fig. 1. This figure shows a simplified stereo configuration of two identical cameras with parallel optical axes and with a distance between the optical centers of  $|\mathbf{o}^l - \mathbf{o}^r|$ . The image planes are at a distance of  $z_0$  from their optical centers and the object is located at a depth of  $z$ . Shifting the left optical center onto the right one yields the configuration in the right part of Fig. 1. The two projections of the house are shifted copies of each other with distance  $d$ , the *disparity*. From a simple geometric consideration it can be concluded that  $|\mathbf{o}^l - \mathbf{o}^r|/z = d/z_0$ , and hence  $z = z_0 |\mathbf{o}^l - \mathbf{o}^r|/d$ .

For a general camera configuration, the geometric setting is much more complicated. However, the depth increases always with decreasing disparity (see e.g. [1], page 211, and [2], page 87) and the relationship between disparity and depth can be computed using the calibrated matrices of the two cameras. The detailed analysis of these relationships is out of the scope of this paper.

---

\* This work has been supported by DFG Grant FE 583/1-1.



**Fig. 1.** Geometric setting of a stereo system. Left part: projection of 3D object (house) onto the left and right image plane. The left optical center is denoted by  $\mathbf{o}^l$ ; the right optical center is denoted by  $\mathbf{o}^r$ . Right part: moving the left camera configuration onto the right one yields two projections of the house with a certain displacement, the disparity. The distance between the two back-projections into 3D space is given by the distance of the optical centers

## 1.2 Disparity from Phase

The disparity can be estimated by various methods, e.g. correlation based techniques, feature based techniques, and phase based techniques [3, 4]. Phase based approaches are used for disparity estimation for the following reasons:

- Correlation based approaches are of a higher computational complexity.
- Feature based approaches do not yield unique correspondences between a left and a right image patch, since the data is projected onto a subspace.
- Phase based approaches yield sub-pixel accuracy without additional effort.
- Applying a phase based method in an appropriate multi-scale framework ensures to fulfill the constraint of a continuous disparity.

The phase based approach using classical quadrature filters can be found in various publications. In this paper we just refer to the more recent works in [5, 6, 3]. Further references can also be found in [2], Chap. 7.4.

The basic idea of disparity from phase is as follows. The disparity is estimated from the difference of the local phases obtained from quadrature filters applied to both images. In contrast to global shifts of the signal, the local phase is not linear to the shift parameter [7]. Hence, the shift theorem cannot be applied directly to estimate the disparity. However, by approximating the local phase in a first order Taylor series expansion, it is possible to estimate the disparity from the local phase difference and the local frequency.

Assume that the right image  $I^r(\mathbf{x})$  (where  $\mathbf{x} = (x, y)$ ) is obtained from the left one by the disparity map  $\mathbf{d}(\mathbf{x})$ , i.e.,  $I^r(\mathbf{x}) = I^l(\mathbf{x} - \mathbf{d}(\mathbf{x}))$ . The map  $\mathbf{d}(\mathbf{x})$  can be reduced to a scalar function, since the orientation of disparity is given by the camera geometry. For simplicity let us assume that  $\mathbf{d}(\mathbf{x}) = (d(\mathbf{x}), 0)$ , i.e.,

the disparity is constrained to be a horizontal displacement (we will return to this point later). From these assumptions, we conclude that

$$\varphi^r(\mathbf{x}) = \varphi^l(x - d(\mathbf{x}), y) \quad (1)$$

where  $\varphi(\mathbf{x})$  is the local phase obtained from the response of a *horizontal* quadrature filter. Actually, (1) is an approximation which becomes worse with increasing gradient of  $d$ . Since there is no reason to prefer one of the two images, we will use the symmetric form of (1) in the following:  $\varphi^r(x + d(\mathbf{x})/2, y) = \varphi^l(x - d(\mathbf{x})/2, y)$ . The Taylor series expansions in  $\mathbf{x}$  of these two terms read

$$\varphi^{r/l}(x \pm d(\mathbf{x})/2, y) = \varphi^{r/l}(\mathbf{x}) \pm d(\mathbf{x})/2 \partial_x \varphi^{r/l}(\mathbf{x}) + \mathcal{O}(d(\mathbf{x})^2) . \quad (2)$$

The partial derivative which occurs in this series expansion is the local frequency *in the horizontal orientation*. It is very important to note that the local frequency should be evaluated in the same orientation as the quadrature filter (see e.g. [1], page 396) and should not be mixed up with the magnitude of the gradient of the local phase, i.e., the isotropic local frequency (see e.g. [1], page 404). Plugging (2) into the symmetric version of (1) yields

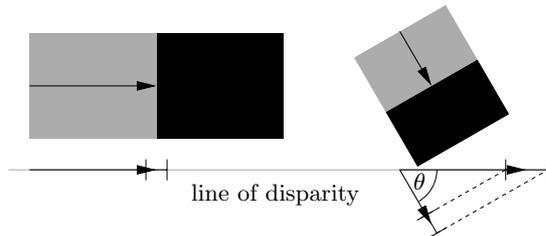
$$d(\mathbf{x}) \approx 2 \frac{\varphi^l(\mathbf{x}) - \varphi^r(\mathbf{x})}{\partial_x \varphi^l(\mathbf{x}) + \partial_x \varphi^r(\mathbf{x})} . \quad (3)$$

Hence, the disparity can be estimated by the quotient of the local phase difference and the mean local frequency in horizontal orientation. If the disparity orientation is not horizontal but given by the unit vector  $\mathbf{e}$ , the corresponding terms in the previous considerations must be changed accordingly. The quadrature filter must be oriented according to  $\mathbf{e}$  and the horizontal derivative is replaced by the directional derivative with respect to  $\mathbf{e}$ .

### 1.3 Reliability of the Estimate

The reliability of the disparity estimate depends on certain signal properties.

1. The approximation (1) becomes worse if the gradient of  $d(\mathbf{x})$  becomes larger. Therefore, it must be assured that the disparity map is smooth to some degree. This will be done by a multi-scale approach, see below.
2. The influence of noise increases if the local amplitude of the filter response is low. If an additive noise model is assumed, the reliability is a linear function of the local amplitude.
3. The reliability depends on the local orientation and the local intrinsic dimension (iD). If the signal is locally 1D (or simple), the signal varies only in one orientation. The reliability of the displacement estimate along this orientation (the *normal displacement*) is independent of the absolute orientation. If the disparity orientation differs from the normal orientation by an angle  $\theta$ , the former is obtained by projecting the displacement onto the disparity line, see Fig. 2. This projection increases the standard deviation of the measurement by  $|\cos(\theta)|^{-1}$ . Accordingly, the reliability is proportional to  $|\cos(\theta)|$  which should be reflected by the measurement.



**Fig. 2.** Reliability of disparity estimates for different normal displacements. The standard deviation, indicated as small intervals at the tips of the displacement vectors, is increased if the displacement is projected onto the disparity line

The quadrature filters which are used in the literature so far are mostly Gabor filters (e.g. [7]) or polar separable quadrature filter (e.g. [5]). The former are clearly sub-optimal with respect to their angular behavior. Gabor filters only have a DC component close to zero if their standard deviation in the frequency domain is small compared to their center frequency. This implies that simple structures with orientations not being identical to the disparity orientation yield too little reliabilities for their disparity estimates. Furthermore, the orientation dependency differs with the local frequency. Polar separable filters according to [2] do not show this dependency on the local frequency. However, they also punish the orientation deviation too much ( $\cos(\theta)^2$  instead of  $|\cos(\theta)|$ ).

Hence, the optimal quadrature filter is a polar separable filter with an angular component of  $|\cos(\theta)|$ . However, any filter with an orientation dependent amplitude response mixes up the uncertainty of the measurement caused by noise (second point) and caused by the local orientation (third point). In order to obtain an optimal disparity estimate, it is necessary to keep the influence of noise and of the orientation separated. This is not possible with a classical quadrature filter but it is straightforward with a spherical quadrature filter (SQF) [8].

## 2 The New Approach

An SQF is the *monogenic signal* [9] of a radial bandpass filter. The monogenic signal and its phase approach have already been applied in other applications, see e.g. [10]. The monogenic signal generalizes the analytic signal to 2D by replacing the Hilbert transform with the Riesz transform. The latter is a vector valued LSI operator with the frequency response

$$\mathbf{H}(\mathbf{u}) = i\mathbf{u}/|\mathbf{u}| = i(\cos \psi, \sin \psi) \quad (4)$$

where  $\mathbf{u} = q(\cos \psi, \sin \psi)$  is the frequency vector. Combining a signal  $f(\mathbf{x})$  with its Riesz transform  $\mathbf{f}_R(\mathbf{x}) = (f_{R1}(\mathbf{x}), f_{R2}(\mathbf{x}))$  yields a 3D-vector valued signal, the monogenic signal.

## 2.1 The Geometry of the Monogenic Phase

In [9] it has been shown that the appropriate phase approach for the monogenic signal is a vector valued phase. The phase vector is obtained as

$$\mathbf{r}(\mathbf{x}) = \frac{\mathbf{f}_R(\mathbf{x})}{|\mathbf{f}_R(\mathbf{x})|} \arctan\left(\frac{|\mathbf{f}_R(\mathbf{x})|}{f(\mathbf{x})}\right). \quad (5)$$

According to [9] the *monogenic phase*  $\mathbf{r}$  is identical to the classical phase multiplied by the orientation vector if the underlying signal is i1D. Hence, the local monogenic phase points in a direction perpendicular to a line or an edge.<sup>1</sup>

Under the presence of global shifts, the monogenic phase is linear to displacement *vectors* and not to their absolute values. It is thus straightforward to use the monogenic phase to estimate the normal displacements in a stereo image pair, generalizing the disparity estimation from local phase and local frequency. Assuming that the underlying signal is locally i1D, the normal displacement and the local phase vector are parallel, e.g., both have the orientation  $\mathbf{e}$ . The Taylor series expansion of  $\mathbf{r}(\mathbf{x} \pm \mathbf{d}(\mathbf{x})/2) = \varphi(\mathbf{x} \pm \mathbf{d}(\mathbf{x})\mathbf{e}/2)\mathbf{e}$  reads

$$\mathbf{r}\left(\mathbf{x} \pm \frac{\mathbf{d}(\mathbf{x})}{2}\right) = \mathbf{r}(\mathbf{x}) \pm \frac{\mathbf{d}(\mathbf{x})}{2} \mathbf{e} \cdot \nabla \mathbf{r}(\mathbf{x}) + \mathcal{O}(d^2) = \mathbf{r}(\mathbf{x}) \pm \frac{\mathbf{d}(\mathbf{x})}{2} \nabla \cdot \mathbf{r}(\mathbf{x}) + \mathcal{O}(d^2) \quad (6)$$

where  $\nabla$  is the gradient operator so that  $\mathbf{e} \cdot \nabla$  is the derivative operator in the direction of  $\mathbf{e}$ . Hence, the normal displacement can be estimated by

$$\mathbf{d}(\mathbf{x}) \approx 2 \frac{\mathbf{r}^l(\mathbf{x}) - \mathbf{r}^r(\mathbf{x})}{\nabla \cdot \mathbf{r}^l(\mathbf{x}) + \nabla \cdot \mathbf{r}^r(\mathbf{x})}. \quad (7)$$

Note that the local frequencies  $\nabla \cdot \mathbf{r}(\mathbf{x})$  which occur in this equation are isotropic local frequencies, in contrast to those in (3). These isotropic local frequencies can be estimated by a similar method as described in [1], page 397, by

$$\nabla \cdot \mathbf{r}(\mathbf{x}) = \frac{\mathbf{f}_R(\mathbf{x}) \cdot (\nabla f(\mathbf{x})) - f(\mathbf{x})(\nabla \cdot \mathbf{f}_R(\mathbf{x}))}{f^2(\mathbf{x}) + |\mathbf{f}_R(\mathbf{x})|^2}. \quad (8)$$

In order to avoid phase wrappings in the numerator of (7), the difference of the phase vectors is replaced by

$$\Delta \mathbf{r}(\mathbf{x}) = \frac{f^l(\mathbf{x})\mathbf{f}_R^r(\mathbf{x}) - f^r(\mathbf{x})\mathbf{f}_R^l(\mathbf{x})}{|f^l(\mathbf{x})\mathbf{f}_R^r(\mathbf{x}) - f^r(\mathbf{x})\mathbf{f}_R^l(\mathbf{x})|} \arctan\left(\frac{|f^l(\mathbf{x})\mathbf{f}_R^r(\mathbf{x}) - f^r(\mathbf{x})\mathbf{f}_R^l(\mathbf{x})|}{f^l(\mathbf{x})f^r + \mathbf{f}_R^l(\mathbf{x}) \cdot \mathbf{f}_R^r(\mathbf{x})}\right). \quad (9)$$

## 2.2 Disparity from Monogenic Phase

The reliability of the normal displacement estimate according to (7) is given by the local amplitudes of the two monogenic signals. In order to turn the displacement into a disparity measure, the former must be projected onto the disparity

<sup>1</sup> Whereas in [9] the phase vector is defined such that it points parallel to the structure, (5) lets it point in the perpendicular direction, i.e., it is rotated by  $\pi/2$ .

line as sketched in Fig. 2. Let  $\mathbf{d}'(\mathbf{x})$  denote the disparity along the line given by  $\mathbf{e}$ , this projection is obtained as

$$\mathbf{d}'(\mathbf{x}) = |\mathbf{d}(\mathbf{x})|^2 / (\mathbf{e} \cdot \mathbf{d}(\mathbf{x})) \mathbf{e} , \quad (10)$$

which yields an increase of the standard deviation by  $\cos \theta = (\mathbf{e} \cdot \mathbf{d}) / |\mathbf{d}(\mathbf{x})|$ .

Thus, we have established a formalism for estimating the disparity which keeps track of the uncertainty due to noise and which treats the local geometry in an appropriate way. Furthermore, the disparity estimation from the monogenic phase is also more flexible with respect to the choice of the line of disparity. Whereas the classical method is more or less fixed to horizontal displacements, the new approach is *independent of the line of disparity* unless the projection (10) is performed. This can be used to estimate the actual line of disparity from displacement field. If, in another application, the disparity orientation is not constant but the cameras are calibrated, it is possible to project the displacement vector onto the epipolar line in order to get the disparity. Another possible setting are images from more than two cameras. The displacement vectors can then be combined to compute the most probable depth.

### 3 Implementation

#### 3.1 The Disparity Estimation Algorithm

As stated above, the gradient of the disparity should not be too large for (1) being sufficiently accurate. This can be ensured by implementing the disparity estimation in a multi-scale algorithm. The two images are lowpass filtered and subsampled in order to obtain a resolution pyramid (see e.g. [11, 12]). The disparity estimation starts at the coarsest scale, giving a coarse disparity map. This is then used to compensate large displacements on the next finer level. The modified images on this level are used to refine the disparity map. This procedure is repeated until the finest scale is reached. This multi-scale method, or hierarchical method, is illustrated in e.g. [2].

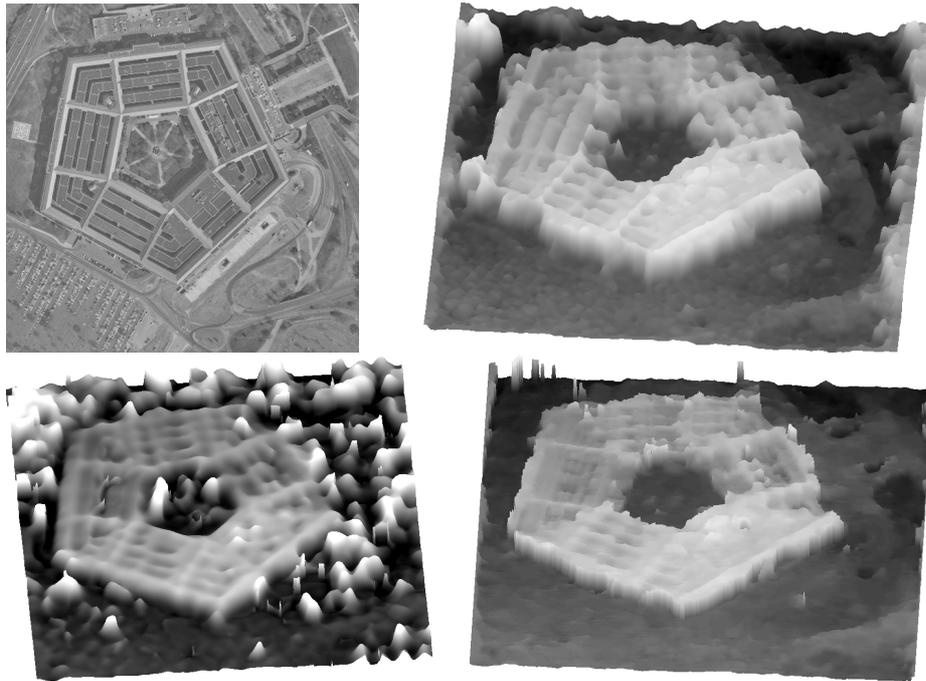
The filters which are applied to estimate the local phase and the local frequency are based on a radial bandpass filter given by  $B(q) = \cos^2((5q - 3\pi)/4)$  if  $q \in [\pi/5, \pi]$  and zero elsewhere. From this frequency response and its Riesz transform a  $9 \times 9$  SQF is obtained by a weighted least square optimization (see e.g. [2]). The partial derivatives of the SQF, which are needed for (8), are obtained by a similar optimization as  $11 \times 11$  filter masks. All optimized filters have a relative error of less than 5%.

Throughout the theoretic part, we have focused on having correct certainty measures. However, it is not trivial to propagate such measures through the multi-scale algorithm. An appropriate tool to represent data *and* its reliability is the *channel representation* [3, 13]. Averaging information in the channel representation is similar to a normalized convolution (see e.g. [14]), i.e., the averaging is weighted by some certainty measure. Furthermore, the channel representation prevents edges from being blurred. Averaging in the channels does not imply

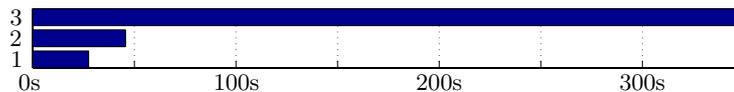
that the decoded information itself will be blurred. The disparity estimates on each scale are stored *and added* as channels. Addition of channels yields a multiplication of the certainty measure, which is reasonable since the measurements are taken from disjunct frequency components and can therefore be considered as independent. In order to combine a coarser disparity map with a finer one, the former has to be interpolated. This is also done in the channel representation which yields a good noise suppression without blurring the disparity map.

### 3.2 Experiment

We have applied the described algorithm to a typical test image pair, showing the Pentagon, see Fig. 3. Compared to the result of a disparity algorithm based on classical quadrature filters and normalized convolution (see [5]), the noise is reduced while the steps in the disparity map are preserved better. Compared to the result from [3], the noise suppression and the preserving of edges are similar. However, the latter disparity map shows some less accurate estimates, see, e.g., the bridge in the upper right corner. Furthermore, the new approach is the fastest one and reduces the complexity by 40% compared to the simple approach based on classical quadrature filters (see Fig. 4).



**Fig. 3.** Upper row: left stereo image (left) and disparity obtained from the presented algorithm (right). Bottom row: disparities obtained from the approaches according to [5] (left) and according to [3] (right)



**Fig. 4.** Run-time comparison of Matlab-implementations on a Sun Ultra 60 for the methods according to [3] (3), according to [5] (2), and according to this paper (1)

## 4 Conclusion

We have presented a new method for disparity estimation, based on the monogenic phase. We have replaced the classical quadrature filters by spherical quadrature filters, motivated by some considerations of the measurement reliability. The resulting estimation formula is more flexible than the one based on horizontal quadrature filters. The implemented algorithm is based on a scale-pyramid and the channel representation in order to propagate the certainties through the hierarchy of the processing. The presented experiment shows that the disparity estimation from monogenic phase is less noise sensitive than the classical method. Furthermore, edges in the resulting disparity maps are preserved as good as for more complex methods like the one based on canonical correlation.

## References

1. Jähne, B.: *Digitale Bildverarbeitung*. Springer-Verlag, Berlin (1997)
2. Granlund, G.H., Knutsson, H.: *Signal Processing for Computer Vision*. Kluwer Academic Publishers, Dordrecht (1995)
3. Borga, M.: *Learning Multidimensional Signal Processing*. PhD thesis, Linköping University, Sweden (1998)
4. Scharstein, D., Szelisky, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Journal of Computer Vision* **47** (2002) 7–42
5. Westelius, C.J.: *Focus of Attention and Gaze Control for Robot Vision*. PhD thesis, Linköping University, Sweden (1995)
6. Hansen, M.: *Stereosehen - ein verhaltensbasierter Ansatz*. PhD thesis, Inst. f. Inf. u. Prakt. Math. der Christian-Albrechts-Universität Kiel (1998)
7. Hansen, M., Daniilidis, K., Sommer, G.: Optimization of stereo disparity estimation using the instantaneous frequency. In: *Proc. Computer Analysis of Images and Patterns*. Volume 1296 of LNCS, Springer-Verlag (1997) 321–328
8. Felsberg, M.: *Low-Level Image Processing with the Structure Multivector*. PhD thesis, Inst. f. Inf. u. Prakt. Math. der Christian-Albrechts-Universität Kiel (2002)
9. Felsberg, M., Sommer, G.: The monogenic signal. *IEEE Transactions on Signal Processing* **49** (2001) 3136–3144
10. Felsberg, M., Sommer, G.: A new extension of linear signal processing for estimating local properties and detecting features. In: *22. DAGM Symposium Mustererkennung*, Springer-Verlag (2000) 195–202
11. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. *IEEE Trans. Communications* **31** (1983) 532–540
12. Granlund, G.H.: In search of a general picture processing operator. *Computer Graphics and Image Processing* **8** (1978) 155–173
13. Nordberg, K., Granlund, G., Knutsson, H.: Representation and Learning of Invariance. In: *Proc. IEEE Int'l Conf. on Image Processing*, Austin, Texas (1994)
14. Forssén, P.E., Granlund, G., Wiklund, J.: Channel representation of colour images. Technical Report LiTH-ISY-R-2418, Dept. EE, Linköping University (2002)