

Improved Confidence Measures for Variational Optical Flow

Maren Brumm, Jan Marek Marcinczak and Rolf-Rainer Grigat

TU Hamburg-Harburg, Harburger Schlossstraße 20, 21079 Hamburg, Germany

Keywords: Variational Optical Flow, Confidence Measure, Performance Evaluation, Structure-Texture Decomposition.

Abstract: In the last decades variational optical flow algorithms have been intensively studied by the computer vision community. However, relatively few effort has been made to obtain robust confidence measures for the estimated flow field. As many applications do not require the whole flow field, it would be helpful to identify the parts of the field where the flow is most accurate. We propose a confidence measure based on the energy functional that is minimized during the optical flow calculation and analyze the performance of different data terms. For evaluation, 7 datasets of the Middlebury benchmark are used. The results show that the accuracy of the flow field can be improved by 53.3 % if points are selected according to the proposed confidence measure. The suggested method leads to an improvement of 35.2 % compared to classical confidence measures.

1 INTRODUCTION

Since Horn and Schunck (Horn and Schunck, 1981) variational optical flow has been an active field of research. An optical flow field gives for every pixel in the first input image a motion vector that estimates the movement to its new position in the second image. There are many different applications for which the optical flow field can be used, such as motion segmentation or 3D reconstruction. However, the optical flow estimation is difficult in areas with changing illumination or complex movements. Commonly, there are areas, where the flow field is less accurate than in others. Many applications require the flow field only for a subset of pixels. Therefore, a confidence measure is needed to identify the parts of the field where the flow is most reliable. Another issue is continuous tracking, where points are tracked over several frames. Here, a confidence measure can be used to detect vanishing points or poorly estimated points that would lead to large errors if tracked on.

Barron et al. (Barron et al., 1994) proposed to use the magnitude of the image gradient as confidence measure. The intention is to reward high structure image parts as the data term has very little informative value in image parts of low structure. However, they demonstrated that this approach is not reliable for variational methods. Bruhn and Weickert argue in (Bruhn and Weickert, 2006) that large gradients commonly result from noise and from pixels that are occluded in the second frame, while areas with small gradients might be filled in accurately by regulariza-

tion.

The approach described in (Bruhn and Weickert, 2006; Bruhn et al., 2005) uses the energy functional that is minimized during the flow calculation as confidence measure. The smaller the contribution of a pixel to the final energy the more reliable it is rated.

In this paper we use a similar approach, but do not limit ourself to using the same data term for the confidence measure as for the optical flow estimation. Instead we analyze the performance of linear and non-linear data terms as well as different constancy assumptions based on a structure-texture decomposition. We especially address the problem of changing illumination by adding a data term that penalizes areas with high illumination changes.

For evaluation seven datasets with ground truth of the Middlebury benchmark (Baker et al., 2011) are used.

Future work will show whether additional improvements can be made by optimizing the weighting between data term and regularization term and by considering different norms on both terms.

2 OPTICAL FLOW ESTIMATION

Let $I: \Omega \rightarrow \mathbb{R}$ with $\Omega \subset \mathbb{R}^2$ be a greyscale image and let (x,y) denote the image coordinates. For an image pair $I_1(x,y)$ and $I_2(x,y)$ the optical flow $\mathbf{u}(x,y) = (u,v)$ describes how the points move between $I_1(x,y)$ and $I_2(x,y)$.

The key assumption for the optical flow estimation is that the grey value of a point in space is constant in all frames. Classically, this constancy assumption is given by (Horn and Schunck, 1981):

$$\frac{\partial I(x, y, t)}{\partial t} = 0. \quad (1)$$

Following the description in (Wedel and Cremers, 2011, p. 9) it can be expressed as follows:

$$I_1(x, y) = I_2(x + u, y + v). \quad (2)$$

Using the first taylor approximation of $I_2(x + u, y + v)$ leads to

$$I_1(x, y) = I_2(x, y) + \nabla I_2(x, y)^T \begin{pmatrix} u \\ v \end{pmatrix}, \quad (3)$$

$$0 = \underbrace{I_2(x, y) - I_1(x, y)}_{I_t} + \nabla I_2(x, y)^T \begin{pmatrix} u \\ v \end{pmatrix}. \quad (4)$$

Denoting the partial derivatives with I_t , I_x and I_y the linearized optical flow constraint reads

$$0 = I_t + I_x u + I_y v. \quad (5)$$

The optical flow constraint usually refers to the constancy of the grey values in the input frames. However, grey values are not illumination invariant. Hence, we consider the use of other pixel properties, see Section 4.

The given constraint leads to an under-determined equation system and cannot be used solely to estimate the optical flow field. Horn and Schunck (Horn and Schunck, 1981) proposed to use the additional assumption that the resulting flow field should be smooth. This leads to the following minimization problem:

$$\min_{u, v} \left\{ \int_{\Omega} |\nabla u| + |\nabla v| + \lambda |I_t + I_x u + I_y v| d\Omega \right\}. \quad (6)$$

The data term enforces the optical flow constraint, while the regularization term favours smooth flow fields by penalizing deviations in the flow field. The parameter λ weights between both terms. Commonly, it is advisable to use the L^1 norm instead of the L^2 norm because it is discontinuity preserving and robust to outliers (Zach et al., 2007).

3 CONFIDENCE MEASURE

The proposed confidence measure is based on the energy functional (6) that is minimized to obtain the optical flow field. As described in (Bruhn and Weickert, 2006) the underlying idea is that the flow field

is most accurate in those areas where the functional has the lowest cost. However, there are different requirements for the calculation of a flow field and the rating of its accuracy. Therefore, we analyze different energy formulations to include additional information which supports classification.

Two different strategies are tested. The first is to use exactly the same energy functional as confidence measure that is used to estimate the optical flow:

$$e_{\text{lin}} = |\nabla u| + |\nabla v| + \lambda \sum_{i=1}^N |I_t^d + I_x^d u + I_y^d v|, \quad (7)$$

where N denotes the number of constancy assumptions used in the data term and $I^d(x, y)$ denotes the pixel property that is assumed to be constant. This implies that the data term is linearized. As the estimated flow is available, there is no need to linearize the data term and the nonlinear data term $|I_1(x, y) - I_2(x + u, y + v)|$ can be calculated directly. Therefore, the following confidence measure provides a higher accuracy as it uses the optical flow constraint directly instead of approximating it:

$$e_{\text{nonlin}} = |\nabla u| + |\nabla v| + \lambda \sum_{i=1}^N |I_1^d(x, y) - I_2^d(x + u, y + v)| \quad (8)$$

Instead of limiting us to using the same constancy assumption in the optical flow constraint for the confidence measure as for the flow calculation, we also consider other constancy assumptions for the confidence measure, as described in the next section.

4 STRUCTURE-TEXTURE DECOMPOSITION

Wedel and Cremer propose in (Wedel and Cremers, 2011, p. 36f) to perform a structure-texture decomposition (Meyer, 2001) on the input images for the calculation of the optical flow field. As shadows show up mainly in the structural part, it increases the robustness to illumination changes to assume the constancy of the textural part only.

However, for the confidence measure both channels carry valuable information. The texture channel provides information on how well a pixel values matches while the structure channel can be used to filter out areas with significant illumination changes where the optical flow is expected to be less reliable. Hence we consider using both the texture and the structure channel for the confidence measure. By decomposing the image and comparing both channels separately additional information can be gained compared to simply comparing the grey values without

decomposition.

The structural part $I_S(\mathbf{x})$ of the grey value image $I(\mathbf{x})$ is obtained by solving

$$\min_{I_S} \int_{\Omega} |\nabla I_S(\mathbf{x})| + \frac{1}{2\theta} (I_S(\mathbf{x}) - I(\mathbf{x}))^2 d\mathbf{x}. \quad (9)$$

The solution to this problem can be found in (Wedel and Cremers, 2011, p. 27f).

The textural part $I_T(\mathbf{x})$ can then be calculated by

$$I_T(\mathbf{x}) = I(\mathbf{x}) - \alpha I_S(\mathbf{x}). \quad (10)$$

Parameters are set according to (Wedel and Cremers, 2011, p. 36f) to $\alpha = 0.95$ and $\theta = 0.125$ and 100 iterations are used to calculate I_S .

5 EVALUATION

To evaluate the performance of the proposed confidence measures, seven datasets of the Middlebury benchmark (Baker et al., 2011) for which a ground truth is available are used.

To measure the accuracy of an estimated optical flow field, the end point error (Otte and Nagel, 1994) is used. It is defined as

$$\epsilon_{\Delta} = \|\mathbf{u} - \mathbf{u}'\|, \quad (11)$$

where \mathbf{u} is the correct displacement vector and \mathbf{u}' the calculated displacement vector.

For evaluation the flow field of each dataset is gradually filtered according to the confidence measure. The average end point error for the remaining flow vectors is calculated in every step. For the calculation of the optical flow field, the texture channel is used in the data term unless stated otherwise. For the energy functional a variety of different data terms is considered. The weighting between regularization term and data term for the confidence measure is set to $\lambda = 0.5$.

A comparison of the efficiency of the confidence measures with a linear and a nonlinear data term for two datasets of the Middlebury benchmark can be seen in Figure 1. For both flow calculation and the confidence measure the texture channel is used in the data term. It can be seen that both confidence measures are suitable to select points with low end point error. For the dataset 'Grove2' the linear and nonlinear confidence measure give similar results for selection rates of 75 % and more, although e_{lin} can reduce the error slightly more. For selection rates below 75 % e_{nonlin} is significantly more effective. It decreases the error monotonically, while for e_{lin} the error rises again for selection rates below 18 %. For the other dataset both confidence measures show a

similar performance. Both monotonically decrease the error. For selection rates below 65 % e_{nonlin} gives better results, for higher selection rates e_{lin} gives better results. Overall e_{nonlin} is more effective, although e_{lin} seems to be more suitable for high selection rates. Taking all selection rates of both datasets into account using the nonlinear data term instead of the linear one improves the results on average by 2.2 %. For a selection rate of 1 % the average improvement is 7.6 %. Therefore, e_{nonlin} is used in the following evaluations.

Figure 2 displays the results of the confidence measure e_{nonlin} on the same datasets for three different data terms in the confidence measure. The different data terms use a constancy assumption on the grey value channel, the texture channel and both the structure and texture channel. It can be seen that the best results are obtained when not only the texture channel of the images is used but the texture and structure channel are used together. The structure channel, which contains most of the shadow and shading, may not be suitable to find a pixel matching itself. Nevertheless, it gives valuable information on how well the pixels are matched because in the areas with significant illumination changes the optical flow is expected to be less reliable. The structure-texture data term also outperforms the grey value data term. So apparently decomposing the image and comparing the channels separately leads to a significant improvement, especially for low selection rates. For the dataset 'Grove2' the end point error can be reduced monotonically from 0.89 px to 0.51 px when using only 1 % of the correspondences. This is an improvement of 43 %. For the dataset 'Hydrangea' the error can even be reduced by 83 %.

The results for five other datasets of the Middlebury benchmark can be seen in Figure 4. The structure-texture data term gives the best results for all datasets, except for 'RubberWhale' where using the texture channel only is most effective for selection rates below 30 %. For higher selection rates the structure-texture data term again gives the best results. The end point error is reduced on average over all datasets evaluated by 53.3 % if 1 % of the flow field is selected. The average improvement over all datasets tested for the structure-texture data term compared to the grey value data term is 11.5 % for a selection rate of 1 % and 3.8 % over all selection rates.

Overall using the nonlinear structure-texture data term improves the performance of the confidence measure by 35.2 % for a selection rate of 1 % compared to using exactly the same data term as for the flow calculation.

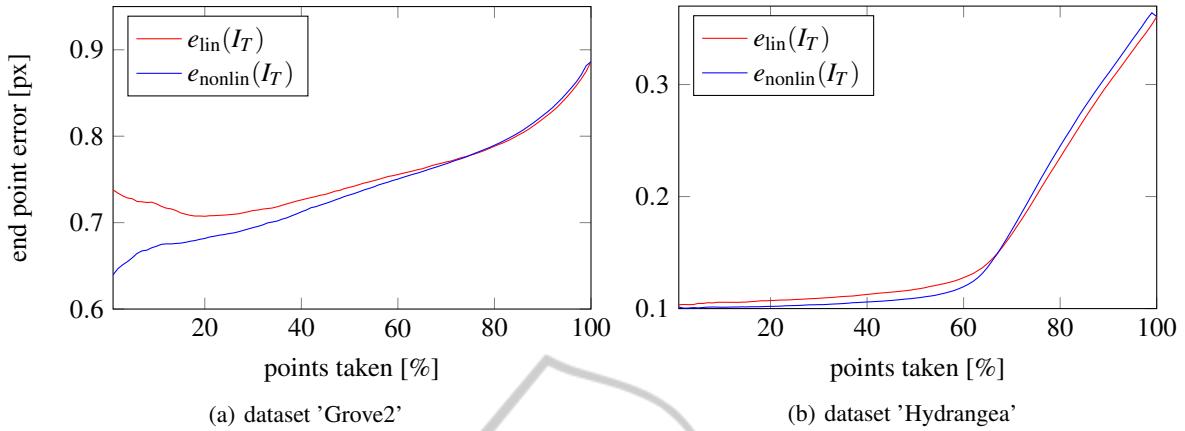


Figure 1: Comparison of the confidence measures with a linear and nonlinear data term for two sequences of the Middlebury benchmark. For selection rates of 65% and lower the nonlinear data term gives better results.

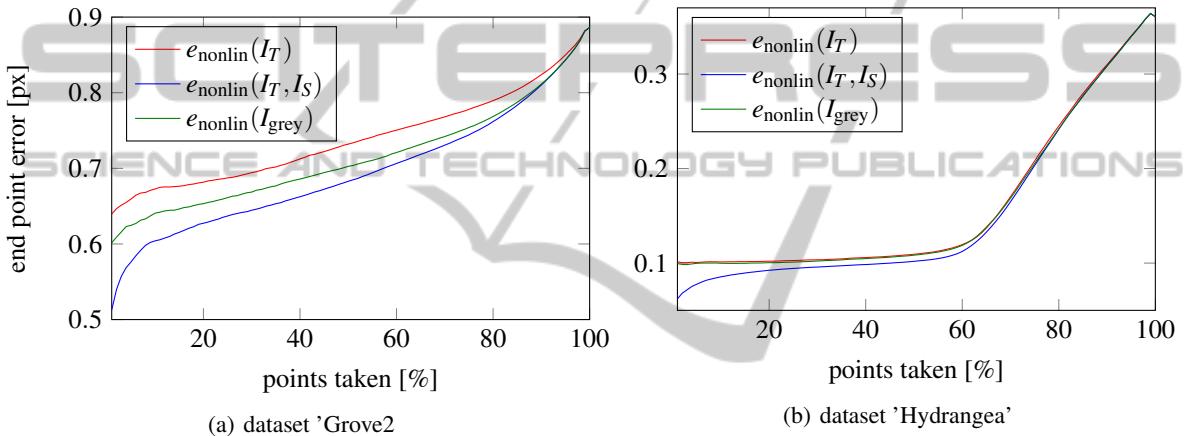


Figure 2: Comparison of different constancy assumptions in the data term. $e(I_T, I_S)$ shows the best results.

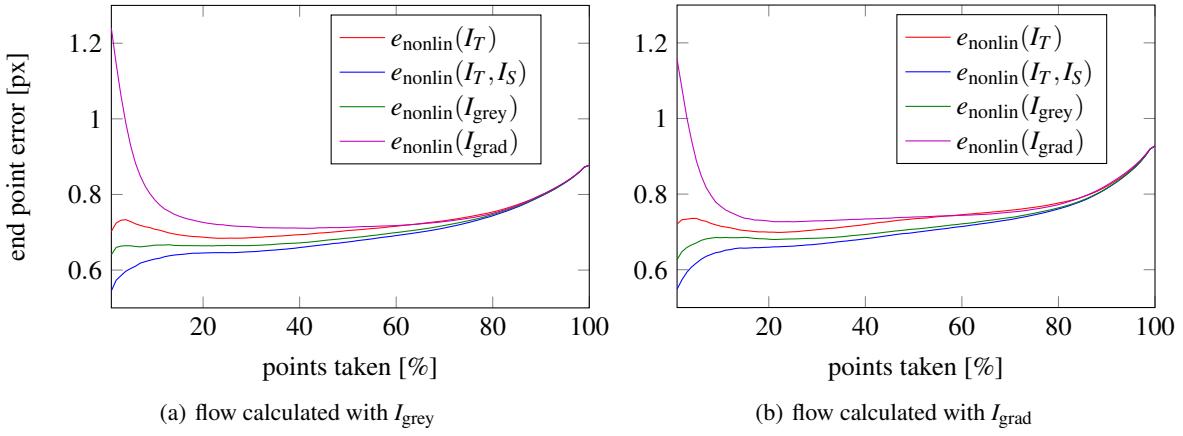


Figure 3: Results for flow fields calculated with different data terms. Again $e(I_T, I_S)$ shows the best results.

To see whether it has an influence which data term is used for the optical flow calculation, the same evaluation is made for the optical flow field calculated with a constancy assumption on the grey values or on the gradient of the image. Here also

the gradient of the images is tested as data term in the confidence measure. The results can be seen in Figure 3. They show that the gradient data term as confidence measure gives worse results than the other data terms. The ranking of the other confidence

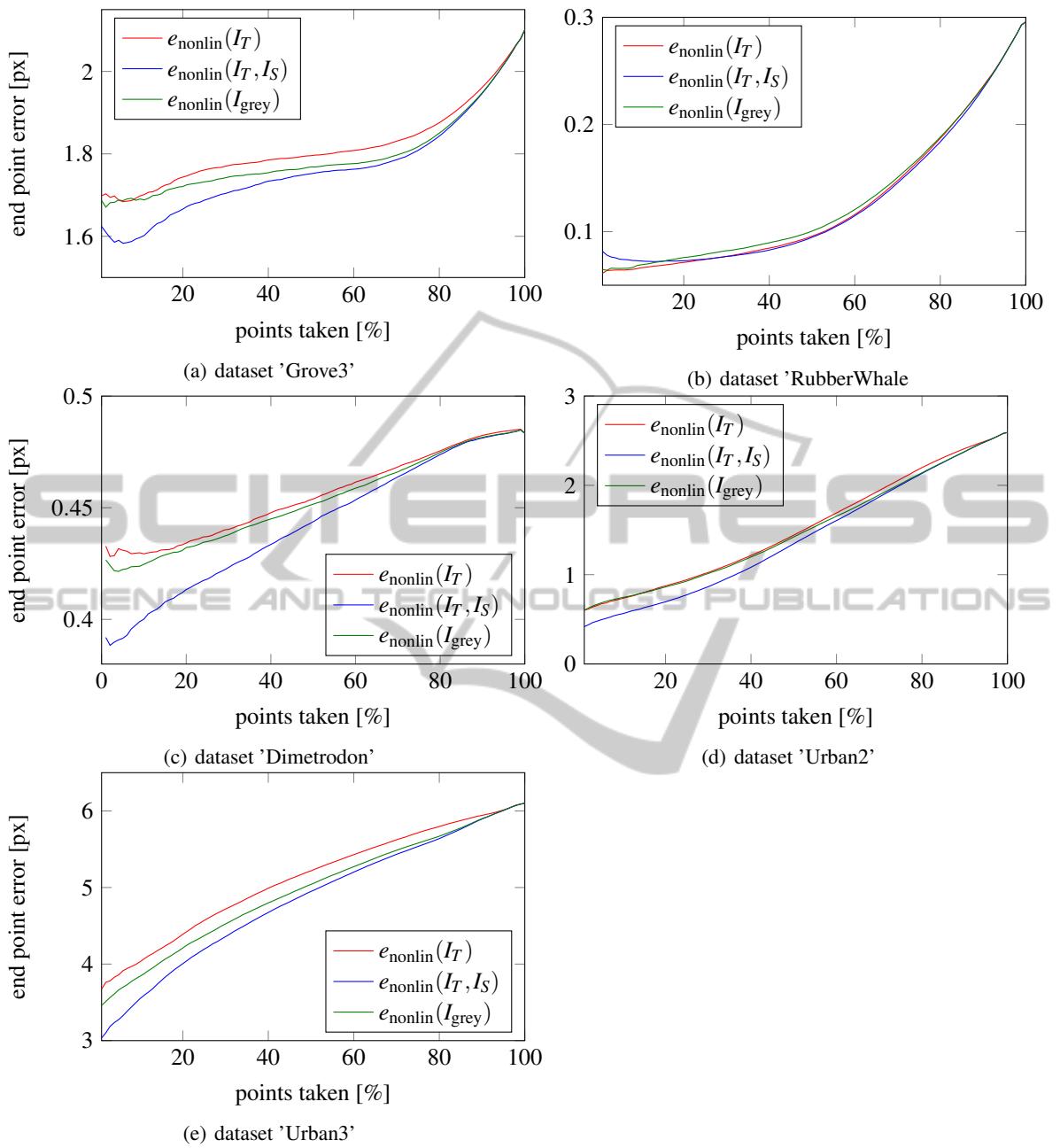


Figure 4: Comparison of different constancy assumptions in the data term for five additional datasets of the Middlebury benchmark.

measures is the same as in the previous experiment and the progression of the error is also similar to the progression in the previous experiment for all three data terms. Therefore, the confidence measure seems to be unaffected by the choice of the data term for optical flow computation.

6 CONCLUSION AND OUTLOOK

To detect areas where the estimated optical flow is most reliable, a confidence measure based on the energy functional that is minimized during the optical flow calculation was evaluated. Both the use of the linear and the nonlinear version of the data term for the confidence measure were considered, as well as

different data terms based on a structure-texture decomposition.

Using the nonlinear data term gives better results for selection rates below 65 %. For higher selection rates the linear data term is more effective, although the difference is not significant.

Decomposing the image in the structure and texture channel of the structure-texture decomposition and using both channels as separate data terms gives the best results for all selection rates on six out of the seven datasets tested. This result seems to be independent on the data term used for flow calculation. The end point error can be reduced by 53.3 % when using only 1 % of the correspondences.

Using the nonlinear structure-texture data term improves the performance of the confidence measure by 35.2 % for a selection rate of 1 % compared to using exactly the same data term as for the flow calculation. Further improvements of the proposed confidence measure may be obtained by optimizing the weighting between the data term and the regularization term. Unlike the energy functional for the flow calculation, the confidence measure does not have to be discontinuity preserving or robust to outliers. On the contrary, it may be eligible to penalize outliers highly. Hence, another possibility would be to use the L^2 norm instead of the L^1 norm on the data term or the regularization term or both.

Wedel, A. and Cremers, D. (2011). *Stereo Scene Flow for 3D Motion Analysis*. Springer.

Zach, C., Pock, T., and Bischof, H. (2007). A duality based approach for realtime $TV-L^1$ optical flow. In *Pattern Recognition*, pages 214–223. Springer.

REFERENCES

- Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J., and Szeliski, R. (2011). A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31.
- Barron, J. L., Fleet, D. J., and Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77.
- Bruhn, A. and Weickert, J. (2006). A confidence measure for variational optic flow methods. In *Geometric Properties for Incomplete Data*, pages 283–298. Springer.
- Bruhn, A., Weickert, J., and Schnörr, C. (2005). Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231.
- Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17:185–203.
- Meyer, Y. (2001). *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, volume 22. American Mathematical Soc.
- Otte, M. and Nagel, H.-H. (1994). Optical flow estimation: advances and comparisons. In *Computer Vision ECCV'94*, pages 49–60. Springer.