# A Statistical Confidence Measure for Optical Flows

Claudia Kondermann<sup>1</sup>, Rudolf Mester<sup>2</sup>, Christoph Garbe<sup>1\*</sup>

<sup>1</sup> Interdisciplinary Center for Scientific Computing, University of Heidelberg, Germany

{claudia.kondermann,christoph.garbe}@iwr.uni-heidelberg.de

Visual Sensorics and Information Processing Lab, University of Frankfurt, Germany mester@vsi.cs.uni-frankfurt.de

Abstract. Confidence measures are crucial to the interpretation of any optical flow measurement. Even though numerous methods for estimating optical flow have been proposed over the last three decades, a sound, universal, and statistically motivated confidence measure for optical flow measurements is still missing. We aim at filling this gap with this contribution, where such a confidence measure is derived, using statistical test theory and measurable statistics of flow fields from the regarded domain. The new confidence measure is computed from merely the results of the optical flow estimator and hence can be applied to any optical flow estimation method, covering the range from local parametric to global variational approaches. Experimental results using state-of-the-art optical flow estimators and various test sequences demonstrate the superiority of the proposed technique compared to existing 'confidence' measures.

#### Introduction 1

It is of utmost importance for any optical flow measurement technique to give a prediction of the quality and reliability of each individual flow vector. This was already asserted in 1994 in the landmark paper by Barron et al. [1], where the authors stated that 'confidence measures are rarely addressed in literature' even though 'they are crucial to the successful use of all [optical flow] techniques'. There are mainly four benefits of confidence measures: 1st) unreliable flow vectors can be identified before they cause harm to subsequent processing steps, 2<sup>nd</sup>) corrupted optical flow regions can be identified and possibly recovered by model-based interpolation (also denoted as 'inpainting'), 3<sup>rd</sup>) existing optical flow methods can be improved, e.g. by integrating the confidence measure into variational approaches, 4<sup>th</sup>) fast, structurally simple optical flow methods in combination with a confidence measure can replace slow, complicated ones. Yet, the confidence measures known today are inadequate for the assessment of the accuracy of optical flow fields due to the following reasons: First, many confidence measures infer confidence values based on the local structure of the image sequence only, without taking into account the computed flow field. Second, most confidence measures are directly derived from specific optical flow computation techniques and, thus, can only be applied to flow fields computed by this method. In fact, so far no generally applicable

<sup>\*</sup> The authors thank the German Research Foundation (DFG) for funding this work within the priority program "Mathematical Methods for Time Series Analysis and Digital Image Processing" (SPP1114).

#### 2 ECCV-08 submission ID 1431

confidence measure exists, which takes into account the computed flow field without being limited to a special type of flow computation method. But if the same model for flow and confidence estimation is used the confidence measure only verifies the restrictions already imposed by the flow computation model. Thus, errors are often not detected as the flow follows the model. Hence, we opt against using the same motion model for confidence estimation. Third, none of the proposed measures is statistically motivated despite the notion 'confidence measure'.

Therefore, in this paper we propose a statistical confidence measure, which is generally applicable independently of the flow computation method. An additional benefit of our method is its adaptability to application-specific data, i.e. it exploits the fact that typical flow fields can be very different for various applications.

## 2 Related Work

The number of previously proposed confidence measures for optical flow fields is limited. In addition to the comparison by Barron et al. [1], another comparison of different confidence measures was carried out by Bainbridge and Lane [2]. In the following we will present confidence measures that have been proposed in the literature so far. Many of these rely on the intrinsic dimensionality of the image sequence. According to [3] the notion 'intrinsic dimension' is defined as follows: 'a data set in n dimensions is said to have an intrinsic dimensionality equal to d if the data lies entirely within a ddimensional subspace'. It has been applied to image processing by Zetzsche and Barth in [4] in order to distinguish between edge-like and corner-like structures in an image. Such information can be used to identify reliable locations, e.g. corners, in an image sequence for optical flow computation, tracking and registration. A continuous formulation has recently been proposed by Felsberg et al. [5]. To make statements on the intrinsic dimension of the image sequence and thus on the reliability of the flow vector, Haussecker and Spies [6] suggested three measures for the local structure tensor method [7]: the temporal coherency measure, the spatial coherency measure and the *corner measure*, which is derived from the two former. All three follow the concept that reliable motion estimation is only possible at those locations in an image sequence where the intrinsic dimension is two, which refers to fully two-dimensional variations in the image plane (e.g. at corners). In case of homogeneous regions and aperture problems, which both correspond to lower intrinsic dimensions, the measures indicate low reliability.

Other examples for confidence measures based on the image structure are the gradient or Hessian of the image sequence or the trace or smallest eigenvalue of the structure tensor [1]. All of these measures are examples for confidence measures which assess the reliability of a given flow vector exclusively based on the input image sequence. In this way they are independent of the flow computation method but they do not take into account the computed flow field.

Other measures take into account the flow field but are derived from and thus limited to special flow computation methods. Examples are the confidence measure proposed by Bruhn and Weickert [8] for variational optical flow methods, which computes the local inverse of the variational energy to identify locations where the energy could not be

minimized, e.g. in cases where the model assumption is not valid. Hence, their approach assigns a low confidence value to these locations. Another example is our previously proposed measure for local optical flow methods [9]. In that paper, the idea is to learn a linear subspace of correct flow fields by means of principal component analysis and then derive the confidence from the distance of the original training flow and the projection into the learned subspace.

Other confidence estimation methods have been suggested by Singh [10], Waxman et al. [11], Anandan [12], Uras et al. [13] as well as by Mester and Hoetter [14]. Yet, these are directly inherent to special optical flow computation methods not applied here.

To obtain a globally applicable, statistically motivated confidence measure which takes into account the flow field we first derive natural motion statistics from sample data and carry out a hypothesis test to obtain confidence values.

Probability distributions for the estimation of optical flow fields have been used before by Simoncelli et al. [15], where the flow is estimated as the solution to a least squares problem. Yet, in their paper the distribution is a conditional based on the uncertainty in the brightness constancy equation. In contrast, we estimate the flow distribution from training data without prior assumptions. The errors in flow estimation have been analyzed by Fermüller et al. [16]. Linear prediction of optical flow by means of motion models has been suggested by Fleet et al. [17].

#### 3 **Natural Motion Statistics**

In order to draw conclusions on the accuracy of a flow vector, we examine the surrounding flow field patch (see Figure 1) of a predefined size  $(n \times n \times T)$ , where  $n \times n$  stands for spatial and T for temporal size). To obtain statistical information on the accuracy, we learn a probabilistic motion model from training data, which can be ground truth flow fields, synthetic flow fields, computed flow fields or every other flow field that is considered correct. In this way, e.g. even motion boundaries can be included in the model if they occur in the training data. If motion estimation is performed for an application domain where typical motion patterns are known a priori, the training data should of course reflect this. It is even possible to use the flow field for which we want to compute the confidence as training data, i.e. finding outliers in one single data set. This leads to a very general approach, which allows for the incorporation of different levels of prior knowledge. To compute the statistical model, the empirical mean m and covariance C

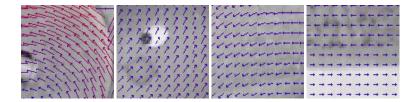


Fig. 1. Examples of flow field patches from which the motion statistics are computed

are computed from the training data set containing the  $n \times n \times T$  flow patches, which are vectorized in lexicographical order. Hence, each training sample vector contains  $p := 2n^2T$  components as it consists of a horizontal and a vertical optical flow component at each patch position. To estimate the accuracy of a given flow vector we carry out a hypothesis test based on the derived statistical model. Note that for results of higher accuracy it is advisable to rotate each training flow field patch four times (each time by 90 degrees) in order to estimate a zero mean vector of the distribution.

# 4 Hypothesis Testing

We want to test the hypothesis

 $H_0$ : 'The central flow vector of a given flow field patch follows the underlying conditional distribution given the remaining flow vectors of the patch.'.

Let D denote the spatio-temporal image domain and  $V : D \to \mathbb{R}^p$  a *p*-dimensional real valued random variable describing possible vectorized flow field patches. Testing the confidence of the central vector of a regarded flow patch boils down to specifying the conditional pdf of the central vector given the remainders of the flow patch, and comparing the candidate flow vector against this prediction, considering a metric induced by the conditional pdf. To define an optimal test statistic, we need to know the correct distribution underlying the flow field patches. Yet, this distribution is unknown. As this is a standard procedure in statistical test theory, we thus choose the optimum test statistic for a reasonable approximation of the conditional pdf. This approximation is here that the conditional pdf of the flow vectors in case that  $H_0$  is true is a two dimensional normal distribution. Even though this approximation is not precisely true, this still leads to a valid test statistic, only the claim that this is the *uniformly most powerful* test statistic is lost. Hence, to develop the test statistic we now assume that V is distributed according to the multivariate normal distribution described by the estimated parameters m and  $\mathbb{C}$ 

$$V \sim \mathcal{N}(\boldsymbol{m}, \mathbf{C})$$
 (1)

with probability density function  $f : \mathbb{R}^p \to \mathbb{R}$ 

$$f(\boldsymbol{v}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{C}|^{\frac{1}{2}}} \exp(-\frac{1}{2} (\boldsymbol{v} - \boldsymbol{m})^T \mathbf{C}^{-1} (\boldsymbol{v} - \boldsymbol{m})) \quad .$$
(2)

We now derive the conditional distribution for the central vector given the remaining vectors of the patch. For a given image sequence location  $(x, y, t) \in D$  let  $v \in \mathbb{R}^p$  correspond to the vectorized flow field patch centered on this location, and let (i, j), i < j denote the line indices of v corresponding to the horizontal and vertical flow vector component of the central vector of the original patch. We partition v into two disjoint vectors, the central flow vector  $v_a$ , and the 'remainders'  $v_b$  of the regarded flow patch:

5

The mean vector and covariance matrix C are partitioned accordingly:

$$oldsymbol{m} = egin{pmatrix} oldsymbol{m}_a \ oldsymbol{m}_b \ oldsymbol{m}_b$$

Then, the conditional distribution  $p(v_a|v_b)$  is a two-dimensional normal distribution with probability density function  $f_{a|b}$ , mean vector  $m_{a|b}$  and covariance matrix  $C_{a|b}$ .

$$\boldsymbol{m}_{a|b} = \boldsymbol{m}_a + \mathbf{C}_{ab}\mathbf{C}_{bb}^{-1}(\boldsymbol{v}_b - \boldsymbol{m}_b) \tag{4}$$

$$\mathbf{C}_{a|b} = \mathbf{C}_{aa} - \mathbf{C}_{ab} \mathbf{C}_{bb}^{-1} \mathbf{C}_{ba} \quad . \tag{5}$$

We stress that these first and second order moments of the conditional pdf are valid independent of the assumption of a normal distribution. To derive the test statistic let

$$d_M : \mathbb{R}^p \to \mathbb{R}_0^+$$
  
$$d_M(v) = (\boldsymbol{v}_a - \boldsymbol{m}_{a|b})^T \mathbf{C}_{a|b}^{-1} (\boldsymbol{v}_a - \boldsymbol{m}_{a|b})$$
(6)

denote the squared Mahalanobis distance between  $v_a$  and the mean vector  $m_{a|b}$  given the covariance matrix  $C_{a|b}$ . The Mahalanobis distance is the optimal test statistic in case of a normally distributed conditional pdf of the central flow vector. This does not imply that the image data or the flow data are assumed to be normally distributed as well. Even though we do not know the conditional distribution, we choose the squared Mahalanobis distance as test statistic. To carry out a hypothesis test (significance test), we have to determine quantiles of the distribution of the test statistic for the case that the null hypothesis to be tested is known to be true. To this end, we compute the empirical cumulative distribution function  $G : \mathbb{R}_+ \rightarrow [0, 1]$  of the test statistic from training data. We obtain the empirical quantile function

$$G^{-1}:[0,1] \to \mathbb{R}_+ \tag{7}$$

$$G^{-1}(q) = \inf\{x \in \mathbb{R} \mid G(x) \ge q\} \quad . \tag{8}$$

To, finally, examine the validity of  $H_0$  we apply a hypothesis test

$$\varphi_{\alpha} : \mathbb{R}^p \to \{0, 1\} \tag{9}$$

$$\varphi_{\alpha}(v) = \begin{cases} 0 & \text{, if } d_M(v) \le G^{-1}(1-\alpha) \\ 1 & \text{, otherwise} \end{cases}$$
(10)

where  $\varphi_{\alpha}(v) = 1$  indicates the rejection of the hypothesis  $H_0$ . Based on this hypothesis test we would obtain a binary confidence measure instead of a continuous mapping to the interval [0, 1]. Furthermore, it would be inconvenient to recompute the confidence measure each time the significance level  $\alpha$  is modified. Therefore, we propose to use the concept of *p*-values introduced by Fisher [18]. A p-value function  $\Pi$  maps each sample vector to the minimum significance level  $\alpha$  for which the hypothesis would still be rejected, i.e.

$$\Pi : \mathbb{R}^p \to [0, 1]$$

$$\Pi(v) = \inf\{\alpha \in [0, 1] | \varphi_\alpha(v) = 1\} = \inf\{\alpha \in [0, 1] | d_M(v) > G^{-1}(1 - \alpha)\}$$
(11)

Hence, we finally arrive at the following confidence measure

$$c : \mathbb{R}^p \to [0, 1]$$

$$c(v) = \Pi(v) = \inf\{\alpha \in [0, 1] | d_M(v) > G^{-1}(1 - \alpha)\} .$$
(12)

As the computation of the confidence measure in fact reduces to the computation of the mean vector and covariance matrix given in (4), the computation can be carried out efficiently.

#### 5 Results

As there are several test sequences with ground truth data and numerous optical flow computation methods with different parameters each, it is impossible to present an extensive comparison between the proposed and previously known confidence measures. Hence, we will present results for a selection of typically used real and artificial sequences and flow computation methods. Here, we will use the Yosemite, the Marble, the Dimetrodon and the RubberWhale sequence (from the Middlebury database [19]). As optical flow computation methods we use the local structure tensor method [7], the non-linear 2d multiresolution combined local global method (CLG) [20] as well as the methods proposed by Nir [21] and Farnebäck [22]. To quantify the error  $e(x) \in \mathbb{R}$  of a given flow vector at image sequence location  $x \in D$  the endpoint error [19] is used. It is defined by the length of the difference vector between the ground truth flow vector  $g(x) \in \mathbb{R}^2$  and the computed flow vector  $u(x) \in \mathbb{R}^2$ :

$$e(x) := \|g(x) - u(x)\|_2 \tag{13}$$

We compare our approach to several of the confidence measures described in section 2. These are the three measures examining the intrinsic dimension of the image sequence by Haussecker and Spies [6] (*strCt*, *strCs*, *strCc*) [6], the inverse of the energy of the global flow computation method by Bruhn et al. [8] (*inverse energy*) [8], the PCA-based measure by Kondermann et al. (*pcaRecon*) [9] and the image gradient measure (*grad*), which is approximated by central differences. In the following, the approach proposed in this paper will be abbreviated by *pVal*. Note that the inverse of the energy measure is only applicable for variational approaches and has thus not been applied to the flow fields computed by methods other than CLG. The Yosemite flow field by Nir et al. [21] was obtained directly from the authors. Hence, no variational energy is available for the computation of the inverse energy confidence measure.

In order to numerically compare the proposed confidence measure to previously used measures we follow the comparison method suggested by Bruhn et al. in [8] called 'sparsification', which is based on quantile plots. To this end, we remove n% of the flow vectors (indicated on the horizontal axis in the following figures) from the flow field in the order of increasing confidence and compute the average error of the remaining flow field. Hence, removing fraction 0 means that all flow vectors are taken into account, so the value corresponds to the average error over all flow vectors. Removing fraction 1 indicates that all flow vectors have been removed from the flow field yielding average error 0. For some confidence measures, the average error even increases after removing

a certain fraction of the flow field. This is the case if flow vectors with errors below the average error are removed instead of those with the highest errors. As a benchmark, we also calculate an 'optimal confidence'  $c_{opt}$ , which reproduces the correct rank order of the flow vectors in terms of the endpoint error (13) and, thus, indicates the optimal order for the sparsification of the flow field:

$$c_{opt}(x) = 1 - \frac{e(x)}{\max\{e(y)|y \in D\}} \quad .$$
(14)

For the experiments the patch size  $n \times n \times T$  was not optimized but kept constant at  $3 \times 3 \times 1$  for all test sequences. The influence of this parameter is rather negligible as shown in Figure 2. Figure 3 shows that the performance of the confidence measure is also mostly independent of the training data. In case ground truth or similar training data is used the performance is improved, but even particle sequence data yields results close to ground truth data.

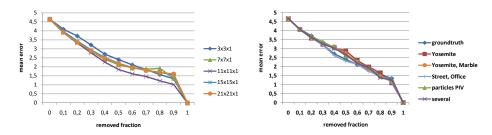


Fig. 2. Remaining mean error for given fraction of removed flow vectors based on different patch sizes for the proposed confidence measure (Farnebäck method on RubberWhale sequence, trained on ground truth data). The results show that the patch size chosen for the confidence measure is rather negligible.

**Fig. 3.** Remaining mean error based on different training sequences for the proposed confidence measure (Farnebäck method on Rubber-Whale sequence for  $3 \times 3 \times 1$  patch size). The results show that the method is hardly sensible to the choice of training data.

Quantile plots of the average flow field error for the state-of-the-art flow computation method by Nir et al. [21], Farnebäck et al. [22], the nonlinear 2d CLG method [20] and the structure tensor method [7] have been computed for the Dimetrodon and the RubberWhale sequence proposed in [19] as well as for the standard Yosemite and Marble test sequences. Selected results are shown in Figures 4 and 5.

For all test examples except for one case the results indicate that the remaining average error for almost all fractions of removed flow vectors is lowest for our proposed confidence measure. As confidence measures are applied to remove the flow vectors with the highest errors only, the course of the curves is most important for small fractions of removed flow vectors and can in practice be neglected for larger fractions. Hence, the results indicate that our proposed confidence measure outperforms the previously employed measures for locally and globally computed optical flow fields on all our test sequences except on the CLG field for the RubberWhale sequence. In this case the

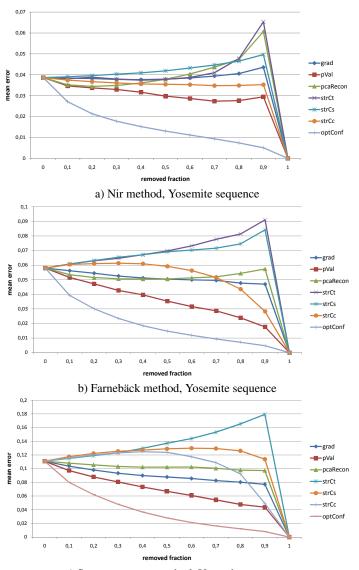
#### 8 ECCV-08 submission ID 1431

inverse energy measure yields results slightly closer to the optimal curve than the proposed measure. It should be noted that for a flow field density of 90% the average error of the local structure tensor method is already lower than that of the CLG flow fields for 100% density on the Marble, Yosemite and RubberWhale test sequences (see for example Figure 5 d),e)). If the CLG flow field is sparsified to 90% as well, the error is approximately equal to that of the structure tensor method for the Yosemite and Marble sequence. Yet, the structure tensor approach only needs a fraction of the computation time of the CLG method and is much simpler to implement. Hence, for the local structure tensor method in two out of three cases we were able to obtain a flow field of 90% density of a quality level equal to that of the CLG method by means of the proposed confidence measure, which clearly shows the benefit of our approach.

To graphically compare confidence measure results we use the structure tensor flow field computed on the RubberWhale test sequence as example as here the difference between the proposed confidence measure and the previously used ones is most eminent. As the scale of confidence measures is not unique we again only compare the order of removal of the flow vectors based on increasing confidence. Hence each flow vector is assigned the time step of its removal from the field. The resulting orders for three of the confidence measures is shown in Figure 6.

### 6 Summary and Conclusion

In this paper we have proposed a confidence measure, which is generally applicable to arbitrarily computed optical flow fields. As the measure is based on the computation of motion statistics from sample data and a hypothesis test, it is to the best of our knowledge the first confidence measure for optical flows, for which the notion 'confidence measure' is in fact justified in a statistical sense. Furthermore, the method can be adapted to specific motion estimation tasks with typical motion patterns by choice of sample data if prior knowledge on the type of computed flow field is available. In this case the results can even be superior to those shown in this paper as here we did not assume any prior knowledge. Results for locally and globally computed flow fields on ground truth test sequences show the superiority of our method compared to previously employed confidence measures. An interesting observation is that by means of the proposed confidence measure we were able to obtain lower average errors for flow fields of 90% density computed by the fast structure tensor method compared to 100 % dense flow fields computed by the non-linear multiresolution combined local global method. And we obtained approximately equal error values for 90% density for both methods. Hence, fast local methods combined with the proposed confidence measure can, in fact, obtain results of a quality equal to global methods, if only a small fraction of flow vectors is removed.



c) Structure tensor method, Yosemite sequence

**Fig. 4.** Average error quantile plot for the comparison of previous confidence measures to the proposed method (pVal); the previous confidence measures are three measures examining the intrinsic dimension of the image (strCt, strCs, strCc) [6], the image gradient (grad), a PCA model based measure [9] (pcaRecon), the inverse of the global energy [8] ( $Inverse\ Energy$ ) and the optimal confidence defined in (14) (optConf); horizontal axis: fraction of removed flow vectors, vertical axis: mean error of remaining flow field.

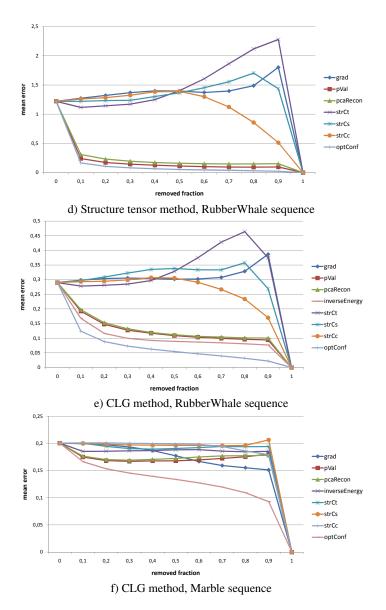
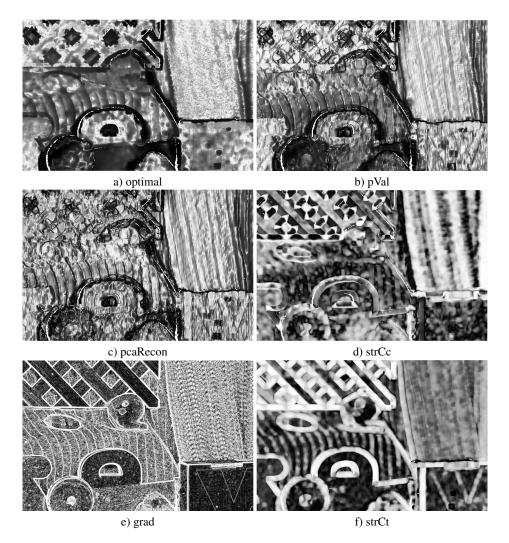


Fig. 5. Average error quantile plots, see Figure 4 for details



**Fig. 6.** Sparsification order of flow vectors based on increasing confidence value for structure tensor flow field on RubberWhale sequence. The proposed confidence measure (pVal) is closest to the optimal confidence.

### References

- Barron, J.L., Fleet, D.J., Beauchemin, S.: Performance of optical flow techniques. International Journal of Computer Vision 12 (1994) 43–77
- Bainbridge-Smith, A. Lane, R.: Measuring confidence in optical flow estimation. IEEE Electronics Letters 32 (1996) 882–884
- 3. Bishop, C. In: Neural Networks for Pattern Recognition. Oxford University Press (1995)
- Zetzsche, C., Barth, E.: Fundamental limits of linear filters in the visual processing of two dimensional signals. Vision Research 30 (1990) 1111–1117
- Felsberg, M., Kalkan, S., Krüger, N.: Continuous dimensionality characterization of image structures. Journal of Image and Vision Computing (2008)
- Haussecker, H., Spies, H.: Motion. In Jähne, B., Haussecker, H., Geißler, P., eds.: Handbook of Computer Vision and Applications. Volume 2. Academic Press (1999) 336–338
- Bigün, J., Granlund, G., J.Wiklund: Multidimensional orientation estimation with applications to texture analysis and optical flow. IEEE Journal of Pattern Analysis and Machine Intelligence 13 (1991) 775–790
- Bruhn, A., Weickert, J.: A confidence measure for variational optic flow methods. In: Geometric Properties for Incomplete data. Springer Netherlands (2006) 283–298
- Kondermann, C., Kondermann, D., Jähne, B., Garbe, C.: An adaptive confidence measure for optical flows based on linear subspace projections. In Hamprect, F., Schnörr, C., Jähne, B., eds.: Pattern Recognition. Volume 4713 of LNCS., Springer (2007) 132–141
- Singh, A.: An estimation-theoretic framework for image-flow computation. In: Proceedings of IEEE ICCV. (1990) 168–177
- Waxman, A., Wu, J., Bergholm, F.: Convected activation profiles and receptive fields for real time measurement of short range visual motion. In: Proceedings of IEEE CVPR. (1988) 717–723
- 12. Anandan, P.: A computational framework and an algorithm for the measurement of visual motion. International Journal of Computer Vision **2** (1989) 283–319
- Uras, S., Girosi, F., Verri, A., Torre, V.: A computational approach to motion perception. Journal of Biological Cybernetics 60 (1988) 79–97
- Mester, R., Hötter, M.: Robust displacement vector estimation including a statistical error analysis. In: Proc. 5th International Conference on Image Processing and its Applications, Edinburgh, UK, Institution of Electrical Engineers (IEE), London (1995) 168–172
- Simoncelli, E.P., Adelson, E.H., Heeger, D.J.: Probability distributions of optical flow. In: Proc Conf on Computer Vision and Pattern Recognition, IEEE Computer Society (1991) 310–315
- Fermüller, C., Shulman, D., Aloimonos, Y.: The statistics of optical flow. Journal of Computer Vision and Image Understanding 82 (2001) 1–32
- Fleet, D.J., Black, M.J., Yacoob, Y., Jepson, A.D.: Design and use of linear models for image motion analysis. International Journal of Computer Vision 36 (2000) 171–193
- 18. Fisher, R.: Statistical Methods for Research Workers. Oliver and Boyd (1925)
- Baker, S., Roth, S., Scharstein, D., Black, M., Lewis, J., Szeliski, R.: A database and evaluation methodology for optical flow. In: Proceedings of the International Conference on Computer Vision. (2007) 1–8
- Bruhn, A., Weickert, J., Schnörr, C.: Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. International Journal of Computer Vision 61 (2005) 211–231
- Nir, T., Bruckstein, A.M., Kimmel, R.: Over-parameterized variational optical flow. International Journal of Computer Vision 76 (2008) 205–216
- 22. Farnebäck, G.: Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field. In: International Conference on Computer Vision, Proceedings. Volume I., Vancouver, Canada (2001) 171–177