MARGINALIZED MAXIMUM A POSTERIORI HYPER-PARAMETER ESTIMATION FOR GLOBAL OPTICAL FLOW TECHNIQUES

Kai Krajsek and Rudolf Mester

Visual Sensorics and Information Processing Lab,
Institute for Computer Science,
J. W. Goethe University,
Robert Mayer Str. 2-4, 60054 Frankfurt am Main, Germany

Abstract. Global optical flow estimation methods contain a regularization parameter (or prior and likelihood hyper-parameters if we consider the statistical point of view) which control the tradeoff between the different constraints on the optical flow field. Although experiments (see e.g. Ng et al. [Ng and Solo(1997)]) indicate the importance of the optimal choice of the hyper-parameters, only little attention has been focused on the optimal choice of these parameters in global motion estimation techniques in literature so far (the authors are only aware of one contribution [Ng and Solo(1997)] which attempts to estimate only the prior hyper-parameter whereas the likelihood hyper-parameter needs to be known). We adapt the marginalized maximum a posteriori (MMAP) estimator proposed in [Mohammad-Djafari(1995)] to simultaneously estimating hyper-parameters and optical flow for global motion estimation techniques. Experiments demonstrate the performance of this optimization technique and show that the choice of the regularization parameter/hyper-parameters is an essential key-point in order to obtain precise motion estimates.

INTRODUCTION

Motion estimation in image sequences is of crucial importance in computer vision as well as in image processing with a wide range of applications spanning from robot navigation over medical image analysis to video compressions. The motion of a single object, i.e. its displacement vector from frame to frame, which is inferred from brightness changes in the image sequence is denoted as the *optical flow* vector. The set of all optical flow vectors is called the optical flow field. Optical flow estimation methods with a prior term that couples all optical flow vectors are usually characterized as global methods. In this contribution, we develop a marginalized maximum a posteriori (MMAP) estimator for simultaneous estimation of hyper-parameters of the likelihood term as well as of the prior term and the optical flow field. The optimal hyper-parameters can be estimated without any prior knowledge or assumption of the current optical flow field, but due to the Bayesian framework, prior knowledge could also be incorporated, if needed. Little attention has been focused on the optimal choice of the regularization parameter in motion estimation techniques in literature so far (the authors are only aware of one contribution [Ng and Solo(1997)]). Some authors even assume that the optimal choice of

the regularization parameter is of minor importance [Bruhn et al.(2005)]. In contradiction to this assumption, obviously is the different value of the regularization parameter which has been proposed by different authors (0.5 in [Horn and Schunck(1981)], 100 in [Barron et al.(1994)]) and the dependency of the motion estimate on the regularization parameter (as demonstrated in [Barron et al.(1994), Ng and Solo(1997)]). But the proposal of a certain value of the regularization parameter is meaningless since its optimal value depends on image statistics, on the image noise statistics as well as on the statistics of the optical flow field [Krajsek(2006)]. Experiments demonstrate the performance of this optimization technique and show that the choice of the regularization parameter is an essential key-point in order to obtain precise motion estimation.

DIFFERENTIAL MOTION ESTIMATION

The general principle behind all differential approaches to motion estimation is that the conservation of some local image characteristics throughout its temporal evolution is reflected in terms of differential-geometric entities on the space-time signal $s(\mathbf{x})$, $\mathbf{x} = (x,y,t)^T$. In its simplest form, the assumed conservation of brightness along the motion trajectory through space-time leads to the well-known brightness constancy constraint equation (BCCE), where $\mathbf{g}$ denotes the gradient of the gray value signal s, $\mathbf{u}_h = (u_x, u_y, 1)^T$ the homogenous form of the direction of motion and $\mathbf{u} = (u_x, u_y)^T$ the optical flow field

$$\mathbf{g}^T \mathbf{u}_h = 0 \ . \tag{1}$$

Since it is fundamentally impossible to solve for $\mathbf{u}_h$ by a single linear equation (*aperture problem*), additional constraints have to be considered. Whereas local methods minimize an error function over a local area $V \subset \mathcal{A}$ assuming a certain motion model in this neighborhood, global methods [Horn and Schunck(1981), Weickert and Schnörr(2001), Black and Anandan(1993)] estimate the optical flow field by minimizing an error functional (or error function if $\mathbf{u}$ is considered on a discrete grid) over the whole region of interest in space-time. The necessary additional constraint is incorporated by a *regularization term* $\rho(\mathbf{u})$ (ρ denotes an operator acting on $\mathbf{u}$) imposing supplementary information on the solution, e.g. the optical flow field should be smooth and should not vary abruptly [Horn and Schunck(1981)]. The *regularization parameter* λ specifies the influence of the regularization term $\rho(\mathbf{u})$ relative to the *data term* $\psi\left(\mathbf{g}^T \mathbf{u}_h\right)$, ($\psi$=real positive function). The optical flow field is estimated by minimizing

$$J(\mathbf{u}) = \int_{\mathcal{A}} \left(\psi\left(\mathbf{g}^T \mathbf{u}_h\right) + \lambda \rho(\mathbf{u}) \right) \mathrm{d}\mathbf{x} \tag{2}$$

with respect to the optical flow field $\mathbf{u}$. Since the introduction of regularization in motion estimation by Horn and Schunk [Horn and Schunck(1981)], a large class of regularization terms has been examined (for an overview we refer to [Weickert and Schnörr(2001)] and to the references therein). But also new data terms have been proposed [Black and Anandan(1993)]. Recently, a combination of global and local constraints (proposed by Bruhn et al. [Bruhn et al.(2005)], which forms the

combined local-global (CLG) method) has been exemplified to increase precision as well as robustness against noise.

BAYESIAN MOTION ESTIMATION

In a Bayesian formulation (see e.g. [Simoncelli et al.(1991)]), the optical flow is estimated via a *probability density function* pdf which connects the observable signal or its gradient with the entity of interest, the optical flow. In order to design such a pdf, we assume a regular grid in space-time considering only signal values and optical flow vectors on the knots of the grid. Since N knots in space-time are isomorphic to the Euclidian space $\mathrm{I\!R}^N$, the signal and the optical flow field can be expressed by a set of vectors $\mathbf{s} \in \mathrm{I\!R}^N$ and $\mathbf{u} \in \mathrm{I\!R}^{2N}$. The approximated gradients $\mathbf{w}$ of the optical flow components $\mathbf{u}$ as well as the approximated gradients $\mathbf{g}$ of the signal components $\mathbf{s}$ can be written in compact matrix vector equations $\mathbf{w} = \mathbf{Hu} \in \mathrm{I\!R}^{6N}$ and $\mathbf{g} = \mathbf{Ps} \in \mathrm{I\!R}^{3N}$. The matrices $\mathbf{P}$ and $\mathbf{H}$ encode the approximation schemes of the gradient computation by finite differences of values at neighborhood positions. In the Bayesian framework, not only the gradients $\mathbf{g} = (\mathbf{g}(\mathbf{x}_1), \mathbf{g}(\mathbf{x}_2), ..., \mathbf{g}(\mathbf{x}_N))$, but also the estimated parameters $\mathbf{u}$ are considered as realizations of random vectors with corresponding pdfs $p(\mathbf{u})$ and $p(\mathbf{g})$, respectively. Prior knowledge about $\mathbf{u}$ is incorporated into the estimation framework via the *prior* pdf $p(\mathbf{u})$. The *maximum a posteriori* (MAP) estimator infers the optical flow field by maximizing the *posterior* pdf $p(\mathbf{u}|\mathbf{g})$ or minimizing its negative logarithm. Using Bayes' law, this leads to

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\{ -\ln(p(\mathbf{g}|\mathbf{u})) - \ln(p(\mathbf{u})) \right\} \ . \tag{3}$$

The term in the bracket on the right side of equ.(3) is denoted as the *objective function $\mathcal{L}$*. For exponential pdfs with partition functions $Z_L(\alpha)$, $Z_p(\beta)$, energies $J_L(\alpha)$ and $J_p(\beta)$ and corresponding hyper-parameters α, β, the objective function becomes

$$\mathcal{L} = J_L(\alpha) + J_p(\beta) + \ln\left(Z_L(\alpha) Z_p(\beta)\right) \ . \tag{4}$$

After discussing the explicit form of the likelihood energy and the prior energy for the case of motion estimation, we develop a method for optimizing the hyper-parameters directly from the observable data. Since all variational methods are equivalent to a corresponding Bayesian formulation [Krajsek and Mester(2006)] and the regularization parameter corresponds to the ratio of the hyper-parameters α, β, our approach allows to optimize the regularization parameters of global optical flow methods in general.

LIKELIHOOD FUNCTIONS AND PRIOR DISTRIBUTIONS FOR MOTION ESTIMATION

In Bayesian estimation, the relation between the observed data $\mathbf{g}$ and the optical flow field $\mathbf{u}$ has to be established by the likelihood function $p(\mathbf{g}|\mathbf{u})$. The error ε_{tj} in the temporal gradient components $\breve{g}_{tj}$ at position j is here approximated as being identical independent noise which follows a Gaussian distribution, $g_{tj} = \breve{g}_{tj} + \varepsilon_{tj}$, whereas the

spatial gradient components $\mathbf{g}_{sj} := (g_{xj}, g_{yj})$ are assumed to be error free. The BCCE $\mathbf{g}_j^T \mathbf{u}_{hj} = \varepsilon_{tj}$ changes accordingly. We obtain the likelihood function $p(\mathbf{g}_t | \mathbf{u}, \mathbf{g}_s)$ by expressing each realization of the random variable ε_{tj} in the joint pdf $p(\varepsilon_t) = \prod_{j=1}^N p(\varepsilon_{tj})$ by the corresponding BCCE

$$p(\mathbf{g}_t | \mathbf{u}, \mathbf{g}_s) = \frac{1}{Z_L(\alpha)} e^{-\alpha \sum_{j=1}^N \psi_1 \left(\| \mathbf{g}_j^T \mathbf{u}_{hj} \|^2 \right)} , \qquad (5)$$

where $\psi_1(x)$ is a positive symmetric function that is $\psi_1(x) = x$ for Gaussian noise. If the optical flow field can be assumed to be constant within spatial neighborhoods and the errors ε_{tj} within each of these regions are i.i.d Gaussian noise, the likelihood function can be derived as an expression depending on the structure tensors $\mathbf{C}_{gj}$ with the energy function $J_L = \alpha \sum_{j=1}^N \mathbf{u}_{hj}^T \mathbf{C}_{gj} \mathbf{u}_{hj}$. The robust version of the likelihood energy function reads $J_L = \alpha \sum_{j=1}^N \psi_1 \left(\mathbf{u}_{hj}^T \mathbf{C}_{gj} \mathbf{u}_{hj} \right)$. The prior pdf encodes our prior information/assumption of the optical flow field. The prior pdf corresponding to the smoothness assumption reads

$$p(\mathbf{u}) = \frac{1}{Z_p(\beta)} e^{-\beta \sum_{j=1}^N \psi_2 \left(\| \mathbf{w}_j^2 \|^2 \right)} , \qquad (6)$$

where ψ_2 is again a positive symmetric function.

MARGINALIZED MAXIMUM A POSTERIORI HYPER-PARAMETER ESTIMATION

The advantage of the Bayesian formulation of the motion estimation problem is that not only the optical flow but also hyper-parameters can be included into the estimation procedure. Firstly, introduced by MacKay [MacKay(1992)] in the context of interpolation, the Bayesian hyper-parameter estimation techniques have been applied to different kind of problems. Other techniques for hyper-parameter estimation have been developed as well but as mentioned in the introduction, only one [Ng and Solo(1997)] of these has been derived for motion estimation so far. The drawback of the method of Ng et al. is its computational cost: a full search in parameter space is necessary to obtain the optimal regularization parameter. Furthermore, only one hyper-parameter can be estimated which corresponds to the prior hyper-parameter in our approach. Thus, it is necessary to know or estimate otherwise the hyper-parameter α which depends on the noise of the gradient field [Krajsek and Mester(2006)]. On the contrary, our approach allows the estimation of all hyper-parameters directly from the observable data. However, if the noise distribution is known, α can be computed [Krajsek and Mester(2006)] and it is only necessary to estimate β. Based on the approach of Mohammad-Djafari [Mohammad-Djafari(1995)], we derive a *marginalized maximum a posteriori* (MMAP) hyper-parameter estimator for the case of motion estimation which estimates hyper-parameters and the optical flow field simultaneously. The main idea of our approach is to approximate the likelihood function $p(\mathbf{g} | \mathbf{u}, \alpha)$ as well as the prior pdf $p(\mathbf{u} | \beta)$ by Gaussian distributions (if they are not already Gaussian) using a Taylor expansion with

respect to $\mathbf{u}$ up to second order of the logarithm of the corresponding pdf. The joint pdf $\tilde{p}(\mathbf{u}, \mathbf{g}|\alpha, \beta)$ of the gradient field and the optical flow field is then obtained by the product of the approximated prior pdf $\tilde{p}(\mathbf{u}|\beta)$ and the approximated likelihood function $\tilde{p}(\mathbf{g}|\mathbf{u}, \alpha)$

$$\tilde{p}(\mathbf{u}, \mathbf{g}|\alpha, \beta) = \tilde{p}(\mathbf{g}|\mathbf{u}, \alpha)\tilde{p}(\mathbf{u}|\beta) \ , \tag{7}$$

which is again a Gaussian function with respect to the optical flow $\mathbf{u}$ and therefore analytically integrable. Furthermore, the partition function $\tilde{Z}(\alpha, \beta) = \tilde{Z}_L(\alpha)\tilde{Z}_p(\beta)$ of the joint pdf is the product of the partition functions of the Gaussian likelihood function and Gaussian prior pdf and thus also analytically tractable. In the current approximation, the likelihood energy J_L as well as the prior energy J_p are expanded around their individual minimum, which are not known a priori. In order to obtain an approximated energy of the joint pdf around its minimum, we exchange the energy by a Taylor series of the original, non approximated joint pdf, up to second order around its minimum with respect to the optical flow field. The optical flow $\hat{\mathbf{u}}$ maximizing the joint pdf is in fact the entity we are searching for and also not known in advance, but as shown below, can be estimated iteratively. Note that the joint pdf is proportional to the posterior pdf of the optical flow field $\mathbf{u}$ and thus maximizing the joint pdf is equivalent to maximizing the joint pdf with respect to the optical flow field. Integrating the resulting approximated joint pdf $\tilde{p}(\mathbf{u}, \mathbf{g}|\alpha, \beta, \hat{\mathbf{u}})$ over $\mathbf{u}$ results in the likelihood function of the hyper-parameters

$$\tilde{p}(\mathbf{g}|\alpha, \beta, \hat{\mathbf{u}}) = \int \tilde{p}(\mathbf{u}, \mathbf{g}|\alpha, \beta, \hat{\mathbf{u}})d\mathbf{u} \ . \tag{8}$$

If prior knowledge about the hyper-parameters are available, it can be encoded into the corresponding hyper-parameter prior pdfs $p(\alpha)$ and $p(\beta)$. The resulting posterior pdf of the hyper-parameters yields

$$\tilde{p}(\alpha, \beta|\mathbf{g}, \hat{\mathbf{u}}) \propto \tilde{p}(\mathbf{g}|\alpha, \beta, \hat{\mathbf{u}})p(\alpha)p(\beta) \ . \tag{9}$$

The hyper-parameters are then estimated by minimizing the negative logarithm of the posterior pdf $\tilde{p}(\alpha, \beta|\mathbf{g}, \hat{\mathbf{u}})$ with respect to α and β for the present realization of the gradient field $\mathbf{g}$. We now compute the concrete likelihood function of the hyper-parameters. Let $\mathbf{Q}(\hat{\mathbf{u}}, \alpha, \beta)$ denote the Hessian of the joint energy $J(\mathbf{u}, \alpha, \beta) = J_L(\mathbf{u}, \alpha) + J_p(\mathbf{u}, \beta)$ and $\mathbf{A}(\hat{\mathbf{u}}, \alpha), \mathbf{B}(\hat{\mathbf{u}}, \beta)$ the Hessians of the likelihood and prior energy, respectively, taken at the optical flow field for which the joint pdf attains its maximum. The optical flow field which minimizes $J(\mathbf{u}, \alpha, \beta)$ is in fact the MAP estimator of the optical flow field for fixed α, β. With the notation $\hat{J} = J(\hat{\mathbf{u}}, \alpha, \beta)$ and $\hat{\mathbf{Q}} = \mathbf{Q}(\hat{\mathbf{u}}, \alpha, \beta)$ the approximated posterior energy reads

$$J(\mathbf{u}, \alpha, \beta) \approx \hat{J} + \frac{1}{2}(\mathbf{u} - \hat{\mathbf{u}})^T \hat{\mathbf{Q}}(\mathbf{u} - \hat{\mathbf{u}}) \ . \tag{10}$$

Inserting the approximated posterior pdf in equ.(8) and integrating over $\mathbf{u}$ yields the likelihood function of the hyper-parameters

$$\tilde{p}(\mathbf{g}|\alpha, \beta, \hat{\mathbf{u}}) = \frac{(2\pi)^N}{\tilde{Z}_L(\alpha)\tilde{Z}_p(\beta)\left|\hat{\mathbf{Q}}\right|^{\frac{1}{2}}} \exp\left(-\hat{J}\right) \ . \tag{11}$$

Since the computation of the determinant $|\mathbf{Q}(\hat{\mathbf{u}}, \alpha, \beta)|$ is not feasible for usual image sequence sizes, an approximation has to be performed. For computing $|\mathbf{Q}(\hat{\mathbf{u}}, \alpha, \beta)|$, we neglect interactions between different pixels, i.e. the *interaction matrix* $\mathbf{B}(\mathbf{u}, \beta)$ becomes a block diagonal matrix. Then the determinant of $\mathbf{Q}(\hat{\mathbf{u}}, \alpha, \beta)$ factorizes into the product of determinants of $\mathbf{Q}_j(\hat{\mathbf{u}}, \alpha, \beta) = \mathbf{A}_j + \mathbf{B}_j$. The approximated objective function for the hyper-parameters then becomes

$$\mathcal{L}(\hat{\mathbf{u}}, \alpha, \beta) \propto \hat{J} + \frac{1}{2} \sum_{j=1}^{N} \ln\left(\left|\hat{\mathbf{Q}}_j\right|\right) + \ln\left(\tilde{Z}_L \tilde{Z}_p\right) \ . \tag{12}$$

Since $\hat{\mathbf{u}}$ itself depends on the hyper-parameters α, β we have to apply an iterative scheme for estimating $\mathbf{u}, \alpha$ and β simultaneously

$$
\begin{aligned}
\mathbf{u}^{k+1} &= \arg\min_{\mathbf{u}} \left\{ J(\mathbf{u}, \alpha^k, \beta^k) \right\} \\
\alpha^{k+1} &= \arg\min_{\alpha} \left\{ \mathcal{L}(\mathbf{u}^k, \alpha, \beta^k) \right\} \\
\beta^{k+1} &= \arg\min_{\beta} \left\{ \mathcal{L}(\mathbf{u}^k, \alpha^k, \beta) \right\} \ .
\end{aligned}
\tag{13}
$$

Note that the first step in the iterative scheme is nothing but the usual optical flow estimation for fixed hyper-parameters as used usually in global approaches. The second and third term in equ.(12) which distinguishes our objective function from others, enables the simultaneous hyper-parameter and motion estimation.

EXPERIMENTAL RESULTS

In this section, the performance of our MMAP estimator is shown. For the global motion estimation we choose the 3D-linear-CLG estimator as described in [Bruhn et al.(2005)]. For the experiment we used three image sequences, together with their true optical flow [1]: 'Yosemite' (without clouds), 'Diverging Tree' and 'Office'. An averaging volume of size $5 \times 5 \times 5$ was applied with a Gaussian weighting function of width $\sigma = 1$. The derivatives occurring in the BCCE were designed according to [Scharr(2000)] and are of size $5 \times 5 \times 5$. The optical flow $\mathbf{u}$ and the hyper-parameters α and β were simultaneously estimated according to our iterative scheme (13). The first step was performed using the *successive over-relaxation* (SOR) method with a *relaxation* parameter of $\gamma = 1.97$. For the second and third step, we set the derivative of $\mathcal{L}(\hat{\mathbf{u}}, \alpha, \beta)$ with respect to α and β to zero and solve for the corresponding hyper-parameters to obtain a fix-point equation which was solved iteratively. For performance evaluation, the *average angular error* (AAE) [Barron et al.(1994)] was computed. Figure 1 shows one image of the 'Office' sequence (left), the ground truth (middle) and the estimated optical flow field (right). Figure 2 shows the computed AAE depending on the regularization parameter $\lambda = \alpha/\beta$ for the three image sequences.

---

[1] The 'Diverging Tree' sequence has been taken from Barron's web-site, the 'Yosemite' sequence from *"http://www.cs.brown.edu/people/black/images.html"* and the 'Office' sequence from *"http://www.cs.otago.ac.nz/research/vision/"* .
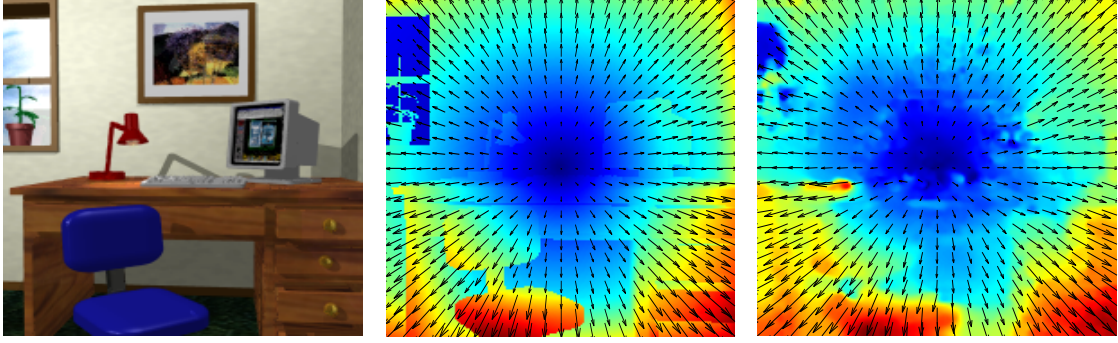
Fig. 1: Left: one image of the 'Office' sequence; middle: ground truth of the 'Office' sequence, the amplitude of the optical flow is encoded in intensity/color with additional optical flow vectors depicted at certain positions; right: estimated optical flow using the 3D-linear-CLG estimator.
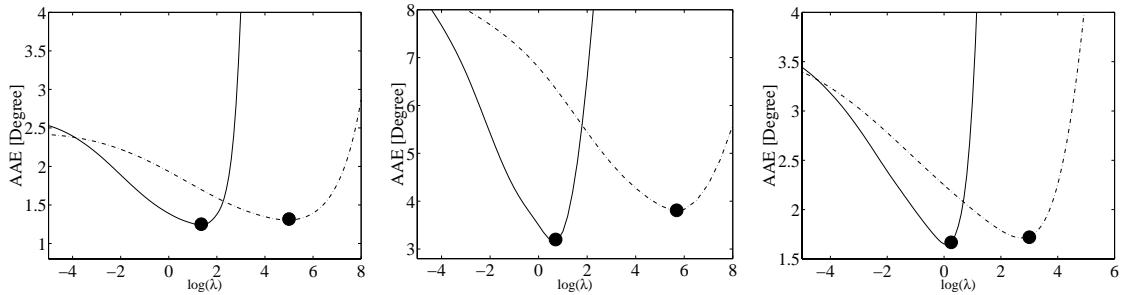

Fig. 2: Average angular error (AAE) vs. regularization parameter $\lambda = \alpha/\beta$ for the image sequences: Diverging Tree (left), Office (middle) and Yosemite (right). The solid lines denote the 3D-nonlinear-CLG estimates whereas the dash-dot lines denote the 3D-linear-CLG estimates. The filled circles denote the corresponding MMAP estimates.

The estimated regularization parameter delivers for all image sequences and all motion estimation techniques a value that is quite close to the optimal value, i.e. the minimum of the corresponding curve. A constant regularization parameter for all sequences as well as for all motion estimation techniques would lead to erroneous results. Thus, λ is a critical parameter which is properly estimated by our MMAP estimator.

SUMMARY AND CONCLUSION

In this contribution, we present a MMAP estimator for simultaneously estimating hyper-parameters and optical flow directly from the observed signal without any prior knowledge of the optical flow. Experiments show that our MMAP estimator delivers the optimal hyper-parameters and show the need for optimizing the hyper-parameters to each image sequence for precise motion estimation. Since there are still free parameters left in global motion estimation techniques like the filter size, our future work will focus on deriving optimization schemes for these still free parameters. Another task will be the incorporation of more complex prior pdfs like those proposed in [Roth and Black(2005)] into our framework.

REFERENCES

[Ng and Solo(1997)] L. Ng, and V. Solo, "A Data-driven Method for Choosing Smoothing Parameters in Optical Flow Problems," in *Proc. International Conference on Image Processing*, Santa Barbara, California, USA, 1997, pp. 360–363.

[Bruhn et al.(2005)] A. Bruhn, J. Weickert, and C. Schnörr, *Int. J. Comput. Vision* **61** (2005).

[Horn and Schunck(1981)] B. Horn, and B. Schunck, *Artificial Intelligence* **17**, 185–204 (1981).

[Barron et al.(1994)] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, *Int. Journal of Computer Vision* **12**, 43–77 (1994).

[Krajsek(2006)] K. Krajsek, A Bayesian framework for differential motion estimation, Technical Report TR-L-0601, Frankfurt University (2006).

[Weickert and Schnörr(2001)] J. Weickert, and C. Schnörr, *Int. J. Comput. Vision* **45**, 245–264 (2001), ISSN 0920-5691.

[Black and Anandan(1993)] M. J. Black, and P. Anandan, "A framework for the robust estimation of optical flow," in *Proc. Fourth International Conf. on Computer Vision, (ICCV93)*, Berlin, Germany, 1993, pp. 231–236.

[Simoncelli et al.(1991)] E. Simoncelli, E. H. Adelson, and D. J. Heeger, "Probability Distribution of Optical Flow," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, 1991, pp. 310–315.

[Krajsek and Mester(2006)] K. Krajsek, and R. Mester, "On the equivalence of variational and statistical differential motion estimation," in *Proc. IEEE SouthWest Symposium on Image Analysis and Interpretation*, 2006, pp. 11–15.

[MacKay(1992)] D. J. C. MacKay, *Neural Computation* **4**, 415–447 (1992).

[Mohammad-Djafari(1995)] A. Mohammad-Djafari, "A full Bayesian approach for inverse problems," in *Proc. 15th International Workshop on Maximum Entropy and Bayesian Methods (MaxEnt95)*, Kluwer, Santa Fe, New Mexico, USA, 1995, pp. 135–144.

[Scharr(2000)] H. Scharr, *Optimal Operators in Digital Image Processing*, Ph.D. thesis, Interdisciplinary Center for Scientific Computing, Univ. of Heidelberg (2000).

[Roth and Black(2005)] S. Roth, and M. J. Black, "On the Spatial Statistics of Optical Flow.," in *ICCV*, 2005, pp. 42–49.