Bayesian Model Selection for Optical Flow Estimation

Kai Krajsek and Rudolf Mester

J.W. Goethe University, Frankfurt, Germany
Visual Sensorics and Information Processing Lab
{krajsek,mester}@vsi.cs.uni-frankfurt.de
http://www.vsi.cs.uni-frankfurt.de

Abstract. Global optical flow techniques minimize a mixture of two terms: a *data term* relating the observable signal with the optical flow, and a *regularization term* imposing prior knowledge/assumptions on the solution. A large number of different data terms have been developed since the first global optical flow estimator proposed by Horn and Schunk [1]. Recently [2], these data terms have been classified with respect to their properties. Thus, for image sequences where certain properties about image as well as motion characteristics are known in advance, the appropriate data term can be chosen from this classification. In this contribution, we deal with the situation where the optimal data term is *not* known in advance. We apply the Bayesian evidence framework for automatically choosing the optimal relative weight between two data terms as well as the regularization term based only on the given input signal.

1 Introduction

Motion estimation in image sequences is of crucial importance in computer vision, it has a wide range of applications spanning from robot navigation over medical image analysis to video compression. The motion of a single object, i.e. its displacement vector from frame to frame, which can be 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. In order to infer the optical flow field from observable entities, e.g. the gray values in an image sequence, a functional relationship between the optical flow field and the observable image signal has to be established. A large number of different types of these *observation equations* has been proposed [2], their properties have been analyzed and classified. It is not very surprising that the simple brightness constancy assumption gives most accurate results when the model assumption - all brightness changes are due to motion - is fulfilled. But if the model assumption is only slightly violated, the accuracy breaks down, leading to highly erroneous results. One way to deal with brightness changes that are not caused by motions is to model the brightness change and optical flow simultaneously [3]. Another way is to relate the optical flow with the signal by observation equations that

F.A. Hamprecht, C. Schnörr, and B. Jähne (Eds.): DAGM 2007, LNCS 4713, pp. 142–151, 2007.

are less sensitive to brightness changes that are not caused by motion [2]. Since they actually disregard parts of the available information, these more robust observation equations have the drawback that they give less accurate results than the simple brightness constancy assumption in case if the model assumption is fulfilled. Furthermore, the most accurate of these illumination-insensitive observation equations induce an orientation dependency such that they are only valid for certain classes of motion [2]. In order to find the best compromise between the different models, it has been proposed to use a combination of different observation equations. But how to find the optimal weight between the different models? The present contribution aims at answering this question which was open so far. It extends the Bayesian evidence framework for choosing the optimal regularization parameter in global optical flow methods presented in [4,5]. Whereas in [4,5] only the optimal weight between data term and regularization term is estimated, the proposed method chooses also the optimal weights between two different model assumptions: the brightness constancy assumption and the generalized constancy assumption that includes the proposed observation equation proposed in [2] as well as the brightness constancy assumption and a new observation equation designed for multiplicative brightness chances.

2 Global Optical Flow Estimation

In the following we describe the image sequence intensity values as a continuous function $s(\boldsymbol{x})$, $\boldsymbol{x} = (x, y, t)$ defined on the continuous Euclidian space denoted as the *space-time volume* $\mathcal{A}$. In order to estimate the optical flow field from the image sequence, a functional relationship, the observation equation, between the signal $s(\boldsymbol{x})$ and the optical flow field $\boldsymbol{u}(\boldsymbol{x})$, has to be established. A simple relation can be derived by the assumption that all intensity variations are due to motion such that the brightness of the signal keeps constant through its evolution in space-time

$$s(x(t), y(t), t) = c \ . \tag{1}$$

This implies the total time derivative to be zero leading to the *brightness constancy constraint equation* (BCCE)

$$g_x u_x + g_y u_y + g_t = 0 \quad \Leftrightarrow \quad \boldsymbol{g}^T \boldsymbol{u}_h = 0 \ , \tag{2}$$

where we have defined $\boldsymbol{g} = (\partial_x s, \partial_y s, \partial_t s)^T$ and $\boldsymbol{u}_h = (u_x, u_y, 1)$. Since it is fundamentally impossible to solve for $\boldsymbol{u}_h$ by a single linear equation (*aperture problem*), additional constraints have to be found and employed. The assumption of spatial [6] or spatiotemporal [7,8] constancy of the flow field $\boldsymbol{u}$ in a local neighborhood V allows the accumulation of all BCCEs in V for a weighted least squares or total least squares optical flow estimation, but this provides the desired disambiguation of optical flow only if the spatial gradients of the image signal vary inside of the regarded neighborhood V. Simoncelli [9] provides a further regularization of the problem by introducing a prior *probability density* *function* (pdf) which penalizes large optical flow vectors. Whereas local methods minimize a loss function (a *residual*) over a local area $V \subset \mathcal{A}$, global methods [1,10,11,12] estimate the optical flow field by minimizing an error functional (or error function if $\boldsymbol{u}$ is considered on a discrete grid) over the whole space-time. The necessary additional constraint is incorporated by a *regularization term* $\rho(\boldsymbol{u})$ (ρ denotes an operator acting on the optical flow $\boldsymbol{u}$) imposing supplementary information on the solution, e.g. the optical flow field should be smooth except for motion boundaries [10]. This means that going from local to global methods is to jump from the simple constant flow assumption directly to expressing smoothness by functionals on derivatives of the resulting flow function. We emphasize here that imposing slightly more complicated local flow models, such as affine, polynomial, etc is still a valid and viable alternative. The *regularization parameter* λ in global approaches specifies the influence of the regularization term $\rho(\boldsymbol{u})$ relative to the *data term* $\psi\left(\boldsymbol{g}^T\boldsymbol{u}_h\right)$, ($\psi$=real symmetric positive function that is monotonically increasing). There is a certain tradition of estimating the optical flow field by minimizing

$$J(\boldsymbol{u}) = \int_{\mathcal{A}} \left(\psi\left(\boldsymbol{g}^T\boldsymbol{u}_h\right) + \lambda\rho(\boldsymbol{u})\right) d\boldsymbol{x} \tag{3}$$

with respect to the optical flow field $\boldsymbol{u}(\boldsymbol{x})$. In principle, the argument of the data term function $\psi(.)$ could be the residual of any valid observation equation. If the brightness constancy assumption does not hold, e.g. due to global brightness changes, (2) does not properly describe the relation between the optical flow and the observable signal any more. One way to deal with this situation is to introduce more complex equations modeling the brightness change and optical flow simultaneously [3]. The drawback is the increase of model parameters that has to be estimated from the input signal. In cases where one is not interested in the brightness model parameters but only in the optical flow it is often more efficient to relate the optical flow with features that are less sensitive to violations of the brightness constancy assumption [2]. A simple and rather popular strategy is to consider the constancy of the spatial gradient of the signal

$$\nabla s(x(t)), y(t), t) = c \ . \tag{4}$$

As in the case of the BCCE, the total time derivative is zero leading to the following equations, denoted as the generalized BCCE (GBCCE) in the following

$$g_{xx}u_x + g_{xy}u_y + g_{xt} = 0 \quad \Leftrightarrow \quad \boldsymbol{g}_x^T\boldsymbol{u}_h = 0 \tag{5}$$

$$g_{yx}u_x + g_{yy}u_y + g_{yt} = 0 \quad \Leftrightarrow \quad \boldsymbol{g}_y^T\boldsymbol{u}_h = 0 \ , \tag{6}$$

where we have defined $\boldsymbol{g}_x = \left(\partial_x^2 s, \partial_x\partial_y s, \partial_x\partial_t s\right)^T$ and $\boldsymbol{g}_y = \left(\partial_y\partial_x s, \partial_y^2 s, \partial_y\partial_t s\right)^T$ respectively. The optical flow is then, as for the case of the brightness constancy assumption, estimated by minimizing the energy functional where we exchange the data term in (3) by $\psi_1\left(\boldsymbol{g}_x^T\boldsymbol{u}_h\right) + \psi_1\left(\boldsymbol{g}_y^T\boldsymbol{u}_h\right)$. The gain in robustness with respect to illumination changes has to be payed with the introduction of directional information in the constancy assumption, i.e. the orientation of the spatial gradients. This means that spatial features are required not to change their orientation through the image sequence, e.g. objects should not to perform a rotation. One way to cope with this limitation is to introduce observation equations based on rotationally invariant features as proposed in [2]. The drawback of this strategy is the apparently poorer performance when compared with the generalized BCCE on an image sequence with violation of the brightness constancy assumption. In [2] a linear combination of different data terms has been proposed. An open question is the choice of the relative weight between both data term and also the choice of the regularization term in this context. This contribution fills this gap by presenting a method for estimating the optimal weights based only on the information delivered by the input signal. In the following we propose alternative illumination change robust feature, the derivative of the logarithm of the signal. Let us assume that the observed signal factorizes into a signal that fulfills the brightness constancy assumption and a term that describes the brightness changes that are not caused by motion. If we consider as a feature the spatial gradient of the logarithm of the signal $\nabla \log f = \nabla \log \gamma + \nabla \log \chi$, the feature separates in the sum of a term that depends on the signal which variations describe the motion and another term that describe all other brightness changes. Taking the total derivative with respect to the time yields the two equations

$$\frac{d\nabla \log f}{dt} = \frac{d\nabla \log \gamma}{dt} + \frac{d\nabla \log \chi}{dt} \quad . \tag{7}$$

If we now assume that χ changes only very slowly in spatial direction, its spatial derivative becomes approximately zero and since per definition $\frac{d\nabla \log \gamma}{dt} = 0$, equations (7) lead to the two linear observation equations

$$h_{xx}u_x + h_{xy}u_y + h_{xt} = 0 \qquad \Leftrightarrow \qquad \boldsymbol{h}_x^T \boldsymbol{u}_h = 0 \tag{8}$$
$$h_{yx}u_x + h_{yy}u_y + h_{yt} = 0 \qquad \Leftrightarrow \qquad \boldsymbol{h}_y^T \boldsymbol{u}_h = 0 \ , \tag{9}$$

where we have defined $h_i = \partial_i \log f$.

In the next section, the variational formulations of the energy functions are reformulated into their statistical equivalent formulation and then the Bayesian evidence framework is presented for estimating the optical flow and model weights simultaneously.

3 Bayesian Motion Estimation

In a Bayesian formulation (see e.g. [9]), the optical flow is estimated via a 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 $\mathbb{R}^N$, the signal and the optical flow field can be expressed by a set of vectors. The gradients $\boldsymbol{w}$ of the optical flow components $\boldsymbol{u}$ as well as the gradients $\boldsymbol{g}$ of the signal components

$\boldsymbol{s}$ can be written in a compact matrix vector equation $\boldsymbol{w} = \mathbf{H}\boldsymbol{u} \in \mathbb{R}^{6N}$, $\boldsymbol{g} = \mathbf{P}\boldsymbol{s} \in \mathbb{R}^{3N}$. In the Bayesian framework, not only the measured gradients $\boldsymbol{g} = (\boldsymbol{g}(\boldsymbol{x}_1), \boldsymbol{g}(\boldsymbol{x}_2), ..., \boldsymbol{g}(\boldsymbol{x}_N))$, but also the estimated parameters $\boldsymbol{u}$ are considered as random variables with corresponding pdfs $p(\boldsymbol{u})$ and $p(\boldsymbol{g})$, respectively. Prior knowledge about $\boldsymbol{u}$ is incorporated into the estimation framework via the *prior* pdf $p(\boldsymbol{u})$. The *maximum a posteriori* (MAP) estimator infers the optical flow field by maximizing the *posterior* pdf $p(\boldsymbol{u}|\boldsymbol{g})$. Using Bayes' law, the posterior pdf can be expressed by the *likelihood function* $p(\boldsymbol{g}|\boldsymbol{u})$, the prior pdf $p(\boldsymbol{u})$ and the gradient pdf $p(\boldsymbol{g})$

$$\hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}} \left\{ \frac{p(\boldsymbol{g}|\boldsymbol{u})p(\boldsymbol{u})}{p(\boldsymbol{g})} \right\} \tag{10}$$

$$= \arg\min_{\boldsymbol{u}} \left\{ -\ln(p(\boldsymbol{g}|\boldsymbol{u})) - \ln(p(\boldsymbol{u})) \right\} . \tag{11}$$

The term in the bracket on the right side of equ.(11) is denoted as the *objective function* $\mathcal{L}$. For Gibbs fields with the partition functions $Z_L(\boldsymbol{\alpha})$, $Z_p(\beta)$, the energies $J_L(\boldsymbol{g}|\boldsymbol{u}, \boldsymbol{\alpha})$ and $J_p(\boldsymbol{u}, \beta)$ and the corresponding hyper-parameters $\boldsymbol{\alpha}$, β, the objective function becomes

$$\mathcal{L} = J_L(\boldsymbol{g}|\boldsymbol{u}, \boldsymbol{\alpha}) + J_p(\boldsymbol{u}, \beta) + \ln\left(Z_L(\boldsymbol{\alpha}) Z_p(\beta)\right) \quad . \tag{12}$$

Note that we parameterize the likelihood energy by multiple hyper-parameters $\boldsymbol{\alpha}$ that weigh different observation models and the prior by one prior hyper-parameter. In the following we describe the likelihood and prior energy for the case of the optical flow estimation. Subsequently, the Bayesian evidence framework for estimating the hyper-parameters is presented.

4 Likelihood Functions and Prior Distributions for Motion Estimation

The likelihood function relates the observable input signal $\boldsymbol{s}$ with the optical flow field $\boldsymbol{u}$. If errors in the spatial gradients can be neglected compared to errors in the temporal gradients, the residuum ε_j of the BCCE's can be assumed to be independent of the optical flow field $\boldsymbol{g}_{sj}^T \boldsymbol{u} + g_{tj} = \varepsilon_j$ [13]. Modeling each random variable ε_j as identical independent distributed, the joint pdf is simply the product $p(\boldsymbol{\varepsilon}_t) = \prod_{j=1}^N p(\varepsilon_{tj})$ of the individual pdfs whereas each pdf is modeled by an exponential distribution. The equations (5), (6), (8) and (9) of the gradient brightness constancy assumption can be reformulated in the same way leading to the corresponding distributions $p(\boldsymbol{\varepsilon}_{xt}) = \prod_{j=1}^N p(\varepsilon_{xtj})$ and $p(\boldsymbol{\varepsilon}_{yt}) = \prod_{j=1}^N p(\varepsilon_{ytj})$, respectively. Due to the linear relationship between the residuum and the temporal gradients g_{tj}, we obtain the following likelihood functions

$$p(\boldsymbol{g}_{kt}|\boldsymbol{u}, \boldsymbol{g}_{ks}, \alpha_k) = \frac{1}{Z_L(\alpha_k)} \exp\left\{ -\alpha_k \sum_{j=1}^{N} \psi_1\left(\boldsymbol{g}_{kj}^T \boldsymbol{u}_{hj}\right) \right\} , \tag{13}$$

where we have introduced $\boldsymbol{g}_1 = \boldsymbol{g}$, $\boldsymbol{g}_2 = \boldsymbol{g}_x$ and $\boldsymbol{g}_3 = \boldsymbol{g}_y$ for notational convenient reasons.If we now assume the different error variables ε_j, ε_{xj} and ε_{yj} to be statistically independent, we can combine all likelihood functions yielding

$$p(\{\boldsymbol{g}_{jt}\}|\boldsymbol{u}, \{\boldsymbol{g}_{js}\}, \{\alpha_j\}) = \prod_{j=1}^{3} p(\boldsymbol{g}_{jt}|\boldsymbol{u}, \boldsymbol{g}_{sj}, \alpha_j) \tag{14}$$

Note that the statistical independence between components of the same gradient is in fact fulfilled, if the temporal gradients are approximated by 1D derivative filter masks. In that case, the error variables of the GBCCE are linear combinations of error variables of the neighborhood of ε_j that do not intersect. In the following, we assume the two hyper-parameters belonging to the GBCCE model are equal reducing the total number of hyper-parameters to α_1 for the BCCE model and α_2 for the GBCCE model.

The prior pdf encodes our prior information/assumption of the optical flow field. The prior pdf corresponding to the smoothness assumption reads

$$p(\boldsymbol{u}) = \frac{1}{Z_p(\beta)} e^{-\beta \sum_{j=1}^{N} \psi_2\left(|\boldsymbol{w}_j|^2\right)} , \tag{15}$$

where ψ_2 is again a positive symmetric function.

5 Bayesian Model Selection

In order to determine the likelihood hyper-parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_L)$ as well as the prior hyper-parameter β, we extend the evidence approach presented in [4,5] from one likelihood hyper-parameter to at least theoretical arbitrary number of likelihood parameters. The reason which allows us for doing so is mainly based on the assumed statistical independence of the likelihood function at different positions in space-time. Note that this is only approximatively true since they are actually correlated due to the overlapping derivative filter masks. We firstly review the main idea of the Bayesian evidence framework before presenting the extensions in more detail. The evidence framework is based on the MAP estimation, i.e. maximizing $p(\boldsymbol{\alpha}, \beta|\boldsymbol{g})$ of the hyper-parameters using the evidence $p(\boldsymbol{g}|\boldsymbol{\alpha}, \beta)$ which is in fact the likelihood function of the hyper-parameters $(\boldsymbol{\alpha}, \beta)$. Multiplying the evidence with the hyper-parameter prior $p(\boldsymbol{\alpha}, \beta)$ yields the joint pdf $p(\boldsymbol{g}, \boldsymbol{\alpha}, \beta)$ of the gradient field and hyper-parameters that is proportional to the desired posterior pdf $p(\boldsymbol{g}|\boldsymbol{\alpha}, \beta)$, i.e. we can obtain the MAP estimate by maximizing also the joint pdf. In the following we assume a constant hyper-parameter prior such that it is sufficient to consider the evidence for estimating the hyper-parameters. The evidence can be obtained from the joint pdf $p(\boldsymbol{u}, \boldsymbol{g}|\boldsymbol{\alpha}, \beta)$ of the gradient field $\boldsymbol{g}$ and the optical flow field $\boldsymbol{u}$ by marginalizing over the optical flow field. The hyper-parameters are then estimated by minimizing the negative logarithm of posterior with respect to $\boldsymbol{\alpha}$ and β for the present realization of the gradient field $\boldsymbol{g}$. In [4] the evidence for the likelihood parameter and the prior hyper-parameters has been derived. In [5] the approach has been extended to two prior hyper-parameters. Following the derivation in [4], we obtain the approximated evidence for multiple likelihood hyper-parameters

$$\tilde{p}(\boldsymbol{g}|\boldsymbol{\alpha}, \beta, \hat{\boldsymbol{u}}) = \frac{(2\pi)^N}{\tilde{Z}_L(\boldsymbol{\alpha})\tilde{Z}_p(\beta)\det \mathbf{Q}^{\frac{1}{2}}} \exp\left(-\hat{J}\right) . \tag{16}$$

where $\hat{\boldsymbol{u}}$ denotes the optical flow field that maximizes the posterior pdf $p(\boldsymbol{u}|\boldsymbol{g}, \boldsymbol{\alpha}, \beta)$ and $\hat{J}$ the energy of the joint pdf $p(\boldsymbol{g}, \boldsymbol{u}|\boldsymbol{\alpha}, \beta)$ taken at $\hat{\boldsymbol{u}}$. The matrix $\mathbf{Q}$ denotes the Hessian of the joint pdf energy $J(\boldsymbol{u}, \boldsymbol{g})$ taken at the maximum of the posterior pdf $p(\boldsymbol{u}|\boldsymbol{g}, \boldsymbol{\alpha}, \beta)$. The partition function are analytically tractable

$$\tilde{Z}_L(\boldsymbol{\alpha}) \propto \prod_j \alpha_j^{-N/2}, \qquad \tilde{Z}_p(\beta) \propto \beta^{-N} \tag{17}$$

due to the Gaussian approximation of prior and likelihood. Note that since we are only interested in the functional dependency on the hyper-parameters, we can get rid of the proportional factors in (17) by maximizing the negative logarithm of the evidence. Since the computation of the determinant $\det \mathbf{Q}$ is not feasible for usual image sequence sizes, a further approximation has to be performed. For computing $\det \mathbf{Q}$, we neglect interactions between different pixels, i.e. $\mathbf{Q}$ becomes block diagonal which is in fact the zero order zone determinant expansion [14] of the matrix $\mathbf{Q}$. Then the determinant of $\mathbf{Q}(\hat{\boldsymbol{u}}, \boldsymbol{\alpha}, \beta)$ factorizes into the product of determinants of $\mathbf{Q}_j(\hat{\boldsymbol{u}}, \boldsymbol{\alpha}, \beta) = \mathbf{A}_j + \mathbf{B}_j$. The approximated objective function for the hyper-parameters then becomes

$$\mathcal{L}(\hat{\boldsymbol{u}}, \boldsymbol{\alpha}, \beta) \propto \hat{J} + \frac{1}{2}\sum_{j=1}^{N} \ln\left(\det \hat{\mathbf{Q}}_j\right) + \frac{N}{2}\sum_{n=1}^{L} \log(\alpha_n) + N\log(\beta) . \tag{18}$$

and the hyper-parameters are estimated by minimizing $\mathcal{L}$. Since $\hat{\boldsymbol{u}}$ itself depends on the hyper-parameters $\boldsymbol{\alpha}, \beta$ we have to apply an iterative scheme for estimating the optical flow field and the hyper-parameters simultaneously, i.e. we estimate the optical flow for fixed hyper-parameters and estimate then the hyper-parameters using the previously estimated optical flow. This procedure is repeated until convergence.

6 Experiments

In this section, the performance of our Bayesian Model selection (BMS) algorithm is presented where we combine either the BCCE (with the likelihood hyper-parameter α_1) with the generalized BCCE (GBCCE) with the spatial gradient of the signal (with the likelihood hyper-parameter α_2) or the spatial gradient of the logarithm of the signal (with the likelihood hyper-parameter α_3). We applied the energy function $\psi_i(x^2) = \xi_i^2\sqrt{1 + x^2/\xi_i^2}$ for all prior and likelihood terms where ξ_i is a free parameter that is to be determined by training data. For the experiment we used three image sequences, together with their

Table 1. Results (expressed by the average angular error (AAE)) of the Bayesian model selection (BMS) experiment with the three image sequences: 'Diverging Tree', 'Yosemite' and 'Office' and the linear combination of two out of three models have been applied. The image sequences fulfill either the brightness constancy assumption, obey a linear decrease of the global brightness with three different gradients $\kappa = 0.05, 0.1, 0.15$ or an exponential brightness decay with three different decay constants $\zeta = 0.025, 0.05, 0.075$.

Diverging Tree	$\kappa = 0$	$\kappa = 0.05$	$\kappa = 0.1$	$\kappa = 0.15$	$\zeta = 0.025$	$\zeta = 0.05$	$\zeta = 0.075$
$\alpha_1 = 1;\ \alpha_2 = 0$	1.11	19.89	38.69	59.51	13.12	22.23	29.39
$\alpha_1 = 0;\ \alpha_2 = 1$	1.56	2.93	5.39	11.59	2.66	3.57	4,43
BMS(α_1, α_2)	**1.28**	**3.46**	**6.62**	**15.72**	**3.23**	**4.29**	**5.57**
$\alpha_1 = 0;\ \alpha_3 = 1$	1.49	1.62	1.82	7.17	1.55	1.83	1.99
BMS(α_1, α_3)	**1.27**	**1.55**	**1.72**	**7.54**	**1.52**	**1.68**	**1.77**
Yosemite							
$\alpha_1 = 1;\ \alpha_2 = 0$	1.84	14.09	31.38	47.25	9.02	17.21	24.73
$\alpha_1 = 0;\ \alpha_2 = 1$	2.07	2.38	3.07	4.38	3.21	3.65	4.15
BMS(α_1, α_2)	**1.72**	**2.24**	**3.04**	**4.66**	**2.17**	**2.56**	**3.02**
$\alpha_1 = 0;\ \alpha_2 = 1$	3.12	3.12	3.12	3.12	3.12	3.12	3.12
BMS(α_1, α_3)	**2.19**	**2.66**	**2.67**	**2.67**	**2.65**	**2.66**	**2.66**
Office							
$\alpha_1 = 1;\ \alpha_2 = 0$	3.28	20.2	31.63	48.54	18.37	26.51	31.77
$\alpha_1 = 0;\ \alpha_2 = 1$	3.77	4.33	5.32	6.87	4.22	4.81	5.41
BMS(α_1, α_2)	**3.21**	**4.43**	**6.17**	**9.37**	**4.29**	**5.23**	**6.36**
$\alpha_1 = 0;\ \alpha_2 = 1$	3.68	3.70	3.72	3.79	3.69	3.70	3.71
BMS(α_1, α_3)	**3.66**	**3.65**	**3.65**	**3.72**	**3.65**	**3.64**	**3.64**

true optical flow [1]: 'Yosemite' (without clouds), 'Diverging Tree' and 'Office'. The derivatives occurring in the BCCE were designed according to [15] and are of size $9 \times 9 \times 9$. The optical flow $\boldsymbol{u}$ and the hyper-parameters α_1, α_2/α_3 and β were simultaneously estimated by minimizing the objective function (18).

For performance evaluation, the *average angular error* (AAE) [16] was computed. We optimized all free parameters, i.e. the pre-smoothing of the image sequences, the parameter ξ_i of the energy functions and the prior hyper-parameter (for cases where the hyper-parameters are not estimated) according to the known ground truth of the 'Diverging Tree' sequence. The algorithm is then applied to the 'Yosemite' and 'Office' sequence with this fixed parameters. We apply the algorithm to the original image sequences and to the image sequences that obey either a global linear brightness change with three different gradients $\kappa = 0.05, 0.1, 0.15$ or an exponential brightness decay with the decay constants $\zeta = 0.025, 0.05, 0.075$. Figure 1 (upper left and upper middle) shows two consecutive frames of the 'Office' image sequence with a linear decrease of brightness with $\kappa = 0.1$. The experimental results are depicted in table 1. Note

[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/"* .

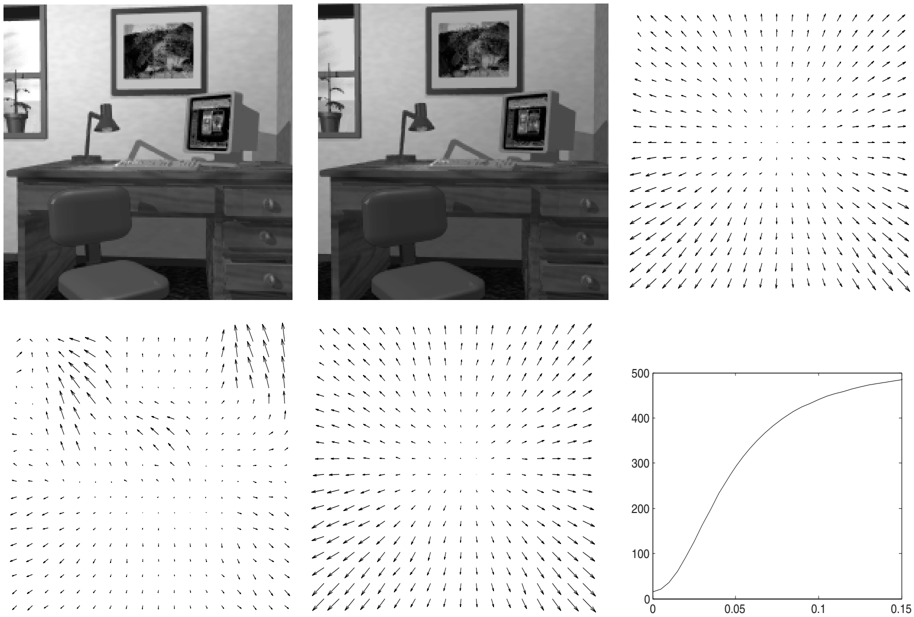

Fig. 1. Upper figures (from left to right): first frame of the 'Office' sequence; second frame of the 'Office sequence' with $\kappa = 0.1$; estimated flow field using $\alpha_1 = 1$ and $\alpha_2 = 0$ for $\kappa = 0$; Lower figures(from left to right):estimated flow field using $\alpha_1 = 1$ and $\alpha_2 = 0$ for $\kappa = 0.1$; estimated flow field using BMS for $\kappa = 0.1$; ratio of both likelihood hyper-parameters α_2/α_1 vs. κ for the BMS algorithm.

that the overall brightness change from one frame to another is rather weak but leads to rather strong erroneous results for the BCCE model ($\alpha_1 = 1$, $\alpha_2 = 0$). The BMS approach gives the most accurate results when applied to the 'test image sequences' 'Yosemite' and 'Office' when compared to fixed models, i.e. only one data term is applied, whose parameters have been tuned to the 'Diverging Tree' sequence . When the brightness constancy assumption is violated the BSE method increases the weight of the second likelihood hyper-parameter α_2 (see figure 1 lower right) in a wide rage depending on the strength of the brightness change leading also to accurate results. Note that in some cases the GBBCE model gives more accurate results than the BMS approach for the sequences with overall brightness changes. But if the optimal model is not known in advance the proposed method estimates automatically the optimal weights between the two models - resulting in most accurate results if the BCCE model is fulfilled and also in most cases in accurate results if the BCCE model assumption is violated.

7 Summary and Conclusion

In this contribution, we presented a Bayesian model selection technique for automatically determining the optimal weights between two data terms in global optical flow methods. We demonstrated the proposed approach with three models: the brightness constancy assumption and gradient brightness constancy assumption of the signal and its logarithm. Further work will examine the expansion of the proposed method to a larger number of models to be selected or weighted. Further research will also focus on the application of the method to the regularization term, i.e. a linear combination of different regularization terms is applied and the optimal weights should optimally be chosen by the Bayesian evidence framework.

References

1. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence 17, 185–204 (1981)
2. Weickert, J., Bruhn, A., Brox, T., Papenberg, N.: A survey on variational optic flow methods for small displacements. In: Scherzer, O. (ed.) Mathematical models for registration and applications to medical imaging, pp. 103–136 (2006)
3. Haussecker, H., Fleet, D.: Computing optical flow with physical models of brightness variation. In: Proc. Computer Vision and Pattern Recognition (2000)
4. Krajsek, K., Mester, R.: Marginalized maximum a posteriori hyper-parameter estimation for global optical flow techniques. In: Bayesian Inference and Maximum Entropy Methods In Science and Engineering, Paris, France (2006)
5. Krajsek, K., Mester, R.: A maximum likelihood estimator for choosing the regularization parameters in global optical flow methods. In: IEEE International Conference on Image Processing, Atlanta, USA (2006)
6. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proc. Seventh International Joint Conference on Artificial Intelligence, Vancouver, Canada, August 1981, pp. 674–679 (1981)
7. Bigün, J., Granlund, G.H.: Optimal orientation detection of linear symmetry. In: Proc. ICCV, pp. 433–438. IEEE, Los Alamitos (1987)
8. Jähne, B.: Digital Image Processing, 4th edn. Springer, Heidelberg (1998)
9. Simoncelli, E., Adelson, E.H., Heeger, D.J.: Probability distribution of optical flow. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, pp. 310–315. IEEE Computer Society Press, Los Alamitos (1991)
10. Weickert, J., Schnörr, C.: A theoretical framework for convex regularizers in pde-based computation of image motion. Int. J. Comput. Vision 45, 245–264 (2001)
11. Black, M.J., Anandan, P.: A framework for the robust estimation of optical flow. In: Proc. Fourth International Conf. on Computer Vision (ICCV93), Berlin, Germany, pp. 231–236 (1993)
12. Alvarez, L., Weickert, J., Sánchez, J.: Reliable estimation of dense optical flow fields with large displacements. Int. J. Comput. Vision 39 (2000)
13. Simoncelli, E.P.: Design of multi-dimensional derivative filters. In: Intern. Conf. on Image Processing, Austin TX (1994)
14. Lee, D.J., Ipsen, I.C.F.: Zone determinant expansions for nuclear lattice simulations. Phys. Rev. C 68, 64003 (2003)
15. Scharr, H.: Optimal Operators in Digital Image Processing. PhD thesis, Interdisciplinary Center for Scientific Computing, Univ. of Heidelberg (2000)
16. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. Int. Journal of Computer Vision 12, 43–77 (1994)