Variational-Bayes Optical Flow

Theodosios Gkamas[1], Giannis Chantas[2] and Christophoros Nikou[1]

Technical Report No 2012-1
Department of Computer Science,
University of Ioannina
Greece

[1]Department of Computer Science, University of Ioannina, Greece.

[2]Department of Informatics, Aristotle University of Thessaloniki, Greece.

Abstract

The Horn-Schunck (HS) optical flow method is widely employed to initialize many motion estimation algorithms. In this work, a variational Bayesian approach of the HS method is presented where the motion vectors are considered to be spatially varying Student's t-distributed unobserved random variables and the only observation available is the temporal image difference. The proposed model takes into account the residual resulting from the linearization of the brightness constancy constraint by Taylor series approximation, which is also assumed to be a spatially varying Student's t-distributed observation noise. To infer the model variables and parameters we recur to variational inference methodology leading to an expectation-maximization (EM) framework with update equations analogous to the Horn-Schunck approach. This is accomplished in a principled probabilistic framework where all of the model parameters are estimated automatically from the data. Experimental results show the improvement obtained by the proposed model which may substitute the standard algorithm in the initialization of more sophisticated optical flow schemes.

Index Terms

Optical flow estimation, variational inference, Bayesian methodology, Student's *t*-distribution.

I. Introduction

The estimation of optical flow is one of the fundamental problems in computer vision as it provides the motion of brightness patterns in an image sequence. This may be useful information, among others, for the determination of 3D structure or the relative motion between the camera and the 3D scene.

Numerous are the proposed methods and their possible categorization in the literature. Nevertheless, not only historically but also from a computational perspective, one may distinguish two main families of methods for optical flow computation. The first category consists of local techniques, relying on an isotropic coarse-to-fine image warping, having as their major representative the Lucas-Kanade algorithm [17]. A Gaussian or rectangular window adapted in scale but being isotropic controls a local neighborhood and jointly with a pyramidal implementation is capable of extending motion estimates from corners to edges and the interior of regions. This method and its variants are still among the most popular for flow and feature tracking. The second family of optical flow methods are the global or variational techniques, relying on an energy minimization framework, with their main representative being the Horn-Schunck method [14], which optimizes a cost function using both brightness constancy and global flow smoothness and has also led to many variants of the basic idea.

The spatial smoothness of the flow field assumed in the above techniques results in many cases to blurred flow boundaries. To overcome this drawback, many researchers proposed various approaches such as robust statistics, treating outliers in both matching and smoothness assumptions [8], [9], [19], variational methodologies [1], [5], [15] incorporating temporal smoothness constraints and gradient constancy assumptions [10], [28], the integration of spatial priors [24], the segmentation of the image pixels or the motion vectors [23], [29], [31]. and learning from ground truth data [25]. Moreover, efforts to combine local and global adaptive techniques were also proposed [13], [26], such as the technique in [11], where the motion vectors are smoothed before being forwarded to a global scheme or the method in [6], where the estimated motion of a feature is influenced by the estimated motion of its neighbors. In both of these methods [11], [6] the spatial integration is isotropic while an anisotropic smoothness term which works complementary with the data term was also conceived [30].

The variational methods belong to the most accurate techniques for optical flow estimation. In this approach, the optical flow is computed as the minimizer of an energy functional consisting of a data term and a smoothness term. The data term is the linearized brightness constancy constraint which results by omitting the higher order terms (by keeping only the first order approximation) of the Taylor series expansion of the constraint. This approximation, which is adopted in order to facilitate the numerical solution is generally not taken into account. However, as it is shown in [1] and [10], this issue should be thoroughly considered.

The smoothness term assumes global or piecewise smoothness spatially. Its properties may vary from homogeneous and isotropic [14], to inhomogeneous [10], or even simultaneously inhomogeneous and anisotropic [21], [27].

Another significant issue in the variational methods is the relative importance between the brightness constancy term and the smoothness term which is usually controlled by a parameter determined by the user remaining fixed during the whole process. This is the case not only for the early algorithm of Horn-Schunck [14] but also for the latest versions of this category of methods [6], [10], [11]. If the weight parameter is not correctly tuned, which is a tedious and prone to errors task for each distinct sequence, it favors one term over the other leading either to motion field degradation or to oversmoothing.

In this paper, we propose a probabilistic formulation of the optical flow problem by following the Bayesian paradigm. The proposed model has intrinsic properties addressing the above mentioned shortcomings.

More specifically, we consider the motion vectors in the horizontal and vertical directions to be independent hidden random variables following a Student's t- distribution. This distribution may model, according to its degrees of freedom, flows following a dominant model (spatial smoothness) as well as flows presenting outliers (abrupt changes in the flow field or edges). Therefore, to account for flow edge preservation with simultaneous smoothing of flat flow regions, the parameter of the t-distribution is also considered to be spatially varying and its value depends on pixel location.

Furthermore, the proposed model takes into account the residual resulting from the linearization of the brightness constancy constraint. The higher order terms of the Taylor series approximation are also represented by a spatially varying Student's t-distributed observation noise. This is, in fact, the only quantity of the model to be considered as observed. By these means, non linear motion changes are also captured.

The form of the assumed distributions makes the marginalization of the complete data likelihood, involving the hidden and the observed quantities, intractable. Thus, to infer the model variables and parameters we recur to variational inference through the mean field approximation [7] which yields a variational expectation-maximization (EM) framework. It turns out that the update solution for the motion field has a form analogous to the update equations of the Horn-Schunck method [14], with the involved quantities being automatically estimated from the two images due to the principled probabilistic modeling. In this framework, we show that the parameter controlling the relative importance of the data and smoothness terms in the standard Horn-Schunck framework is an intrinsic random variable of the proposed model whose statistics are also estimated by the data. Numerical results revealed that the method provides better accuracies not only with respect to standard optical flow algorithms [17], [14] which are used to initialize more sophisticated methods, but also to a recently proposed version of their joint combination [6].

In the remainder of the paper, the modeling of the motion vectors by a t-distribution is presented in section II while the overall probabilistic model for optical flow estimation is described in section III. Model inference is derived in section IV, numerical results are presented in section V and a conclusion is drawn in section VI.

II. A PRIOR FOR THE MOTION VECTORS

Let $\mathbf{I}(\mathbf{x})$ be the first image frame (target frame) containing the intensity values lexicographically and let also $\mathbf{J}(\mathbf{x})$ be the second image frame (source frame) where $\mathbf{x} = (x, y)$ represents the $2D$ coordinates of a pixel. The brightness constancy constraint at a given location is expressed by:

$$\frac{\partial \mathbf{I}}{\partial x}\mathbf{u}_x + \frac{\partial \mathbf{I}}{\partial y}\mathbf{u}_y + \frac{\partial \mathbf{I}}{\partial t} = 0, \tag{1}$$

where we have removed the independent variable representing the location $\mathbf{x}$ for simplicity. In (1), $\mathbf{u}_x$ and $\mathbf{u}_y$ are the motion vectors in the horizontal and vertical directions respectively, $\partial\mathbf{I}/\partial x$ and $\partial\mathbf{I}/\partial y$ are the spatial gradients of the target image and $\partial\mathbf{I}/\partial t$ is the temporal difference between the two images ($\mathbf{J}(\mathbf{x}) - \mathbf{I}(\mathbf{x})$). The above equation holds for any pixel location $\mathbf{x}$ and the determination of the target and source images is a question of convention as they may be interchanged along with a simple sign change.

For convenience, we compactly represent the optical flow values at the i-th location by $\mathbf{u}_k(i)$, for $i = 1, \ldots, N$ where $k \in \{x, y\}$ and N is the number of image pixels. We now assume that $\mathbf{u}_k(i)$ are i.i.d. zero mean Student's *t*-distributed, with parameters λ_k and ν_k:

$$\mathbf{u}_k(i) \sim \mathcal{S}t\left(0, \lambda_k, \nu_k\right), \forall i = 1, ..., N, \forall k \in \{x, y\}. \tag{2}$$

The Student's-*t* distribution implies a two-level generative process [7]. More specifically, $\boldsymbol{\alpha}_k(i)$, $k \in \{x, y\}$ are first drawn from two independent Gamma distributions: $\boldsymbol{\alpha}_k(i) \sim Gamma\left(\frac{\nu_k}{2}, \frac{\nu_k}{2}\right)$. Then, $\mathbf{u}_k(i)$, $k \in \{x, y\}$ are generated from two zero-mean Normal distributions with precision $\lambda_k \boldsymbol{\alpha}_k(i) \mathbf{Q}_i^T \mathbf{Q}_i$ according to $p(\mathbf{u}_k(i)|\boldsymbol{\alpha}_k(i)) = \mathcal{N}(0, (\lambda_k \boldsymbol{\alpha}_k(i) \mathbf{Q}_i^T \mathbf{Q}_i)^{-1})$, where $\mathbf{Q}_i$ is the matrix applying the Laplacian operator to the flow field at the i-th location. Based on the assumption that the flow field should be smooth, it is common to assume this type of prior privileging low frequency motion fields [20], [12].

The probability density function in (2) may be written as

$$p\left(\mathbf{u}_k(i)\right) = \int_0^\infty p\left(\mathbf{u}_k(i)|\boldsymbol{\alpha}_k(i)\right) p\left(\boldsymbol{\alpha}_k(i)\right) d\boldsymbol{\alpha}_k(i), \tag{3}$$

where the variables $\boldsymbol{\alpha}_k(i)$ are hidden because they are not apparent in (2) since they have been integrated out. As the *degrees of freedom* parameter $\nu_k \to \infty$, the pdf of $\boldsymbol{\alpha}_k(i)$ has its mass concentrated around its mean. This in turn reduces the Student's-*t* pdf to a Normal distribution, because all $\mathbf{u}_k(i)$, $k \in \{x, y\}$ are drawn from the same normal distribution with precision λ_k, since $\boldsymbol{\alpha}_k(i) = 1$ in that case. On the other hand, when $\nu_k \to 0$ the prior becomes uninformative. In general, for small values of ν_k the probability mass of the Student's-*t* pdf is more "heavy tailed".

We assume that the horizontal and vertical motion fields are independent at each pixel location. This assumption makes subsequent calculations tractable and is common in Bayesian image analysis.

By defining the $N \times N$ diagonal matrices $\mathbf{A}_k$ = diag$[\boldsymbol{\alpha}_k(1), \ldots, \boldsymbol{\alpha}_k(N)]^T$, $k \in \{x, y\}$, the pdf of the horizontal and vertical motion fields may now be expressed by:

$$p(\mathbf{u}_k|\mathbf{A}_k) = \mathcal{N}\left(\underline{\mathbf{0}}, \left(\lambda_k \mathbf{Q}^T \mathbf{A}_k \mathbf{Q}\right)^{-1}\right), \tag{4}$$

where $\mathbf{Q}$ is the Laplacian operator applied to the whole image and $\underline{\mathbf{0}}$ is a $N \times 1$ vector of zeros. Then, the overall pdf of the motion field $\mathbf{u} = [\mathbf{u}_x, \mathbf{u}_y]^T$ is given by $p(\mathbf{u}) = p(\mathbf{u}_x|\mathbf{A}_x)p(\mathbf{u}_y|\mathbf{A}_y)$, or equivalently:

$$p\left(\mathbf{u}|\tilde{\mathbf{A}}\right) = \mathcal{N}\left(\underline{\mathbf{0}}, \left(\lambda \tilde{\mathbf{Q}}^T \tilde{\mathbf{A}} \tilde{\mathbf{Q}}\right)^{-1}\right), \tag{5}$$

where the $2N \times 1$ vector $\lambda = [\lambda_x, \lambda_y]^T$, the $2N \times 2N$ matrix $\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_y \end{bmatrix}$, the $2N \times 2N$ matrix $\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}$ and $\mathbf{0}$ is a zero matrix of size $N \times N$. Hence, following (5), the marginal distribution $p(\mathbf{u})$ has a closed form.

III. A PROBABILISTIC MODEL FOR OPTICAL FLOW

The optical flow equation (1) may be written in matrix-vector form as:

$$\mathbf{G}\mathbf{u} = \mathbf{d}, \tag{6}$$

where the block diagonal $N \times 2N$ matrix $\mathbf{G} = \begin{bmatrix} \mathbf{G}_x & \mathbf{G}_y \end{bmatrix}$, with $\mathbf{G}_x = \text{diag}\left[\frac{\partial \mathbf{I}(\mathbf{x}_1)}{\partial x}, \ldots, \frac{\partial \mathbf{I}(\mathbf{x}_N)}{\partial x}\right]^T$, $\mathbf{G}_y = \text{diag}\left[\frac{\partial \mathbf{I}(\mathbf{x}_1)}{\partial y}, \ldots, \frac{\partial \mathbf{I}(\mathbf{x}_N)}{\partial y}\right]^T$ contains the spatial derivatives in the horizontal and vertical directions lexicographically and the $N \times 1$ vector $\mathbf{d} = [\mathbf{I}(\mathbf{x}_1) - \mathbf{J}(\mathbf{x}_1), \ldots, \mathbf{I}(\mathbf{x}_N) - \mathbf{J}(\mathbf{x}_N)]^T$ contains the temporal image differences. Therefore, to visually highlight the role of matrix $\mathbf{G}$, eq. (6) may be also written as:

$$\begin{bmatrix} \mathbf{G}_x & \mathbf{G}_y \end{bmatrix} \begin{bmatrix} \mathbf{u}_x \\ \mathbf{u}_y \end{bmatrix} = \mathbf{d}. \tag{7}$$

In order to take into account higher order terms of the Taylor series expansion of the brightness constancy constraint, which are not considered in (1), we add a noise term to (6) yielding:

$$\mathbf{G}\mathbf{u} + \mathbf{w} = \mathbf{d}. \tag{8}$$

We also assume spatially varying Student's *t*-statistics for this $N \times 1$ noise vector:

$$\mathbf{w} \sim \mathcal{N}\left(\underline{\mathbf{0}}, (\lambda_{noise}\mathbf{B})^{-1}\right), \tag{9}$$

where $\lambda_{noise}\mathbf{B}$ is the noise precision matrix and $\mathbf{B} = \text{diag}[\mathbf{b}(1), \ldots, \mathbf{b}(N)]^T$, where the *t*-distribution implies that each $\mathbf{b}(i)$, $i = 1, .., N$ is Gamma distributed with parameter μ:

$$\mathbf{b}(i) \sim Gamma\left(\frac{\mu}{2}, \frac{\mu}{2}\right). \tag{10}$$

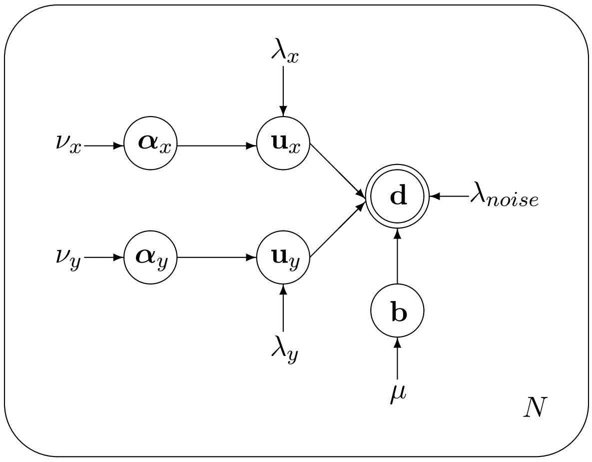

Fig. 1. Graphical model for the optical flow problem.

Following the optical flow matrix-vector formulation in (8) and the noise modeling in (9) and (10), we arrive at the probability of the temporal image differences given the motion vectors:

$$p(\mathbf{d}|\mathbf{u}) = \mathcal{N}\left(\mathbf{G}\mathbf{u}, (\lambda_{noise}\mathbf{B})^{-1}\right). \tag{11}$$

The above probabilistic formulation of the optical flow problem is represented by the graphical model of figure 1. As it may be observed, $\mathbf{d}$ is the vector containing the observations (temporal differences), denoted by the double circle, $\mathbf{u} = [\mathbf{u}_x, \mathbf{u}_y]^T$, $\boldsymbol{\alpha}_x$, $\boldsymbol{\alpha}_y$, $\mathbf{b}$, are the hidden variables of the model, denoted by the simple circles and λ_x, λ_y, λ_{noise}, ν_x, ν_y and μ are the model's parameters. Notice that all of the variables and the observations are of dimension N except of the vector $\mathbf{u}$ collecting the horizontal and vertical motions. This shows the ill-posedness of the original optical flow problem where we seek $2N$ unknowns (vectors $\mathbf{u}_x$ and $\mathbf{u}_y$) with only N observations (vector $\mathbf{d}$).

IV. Model inference

In the fully Bayesian framework, the complete data likelihood, including the hidden variables and the parameters of the model, is given by

$$p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) = p\left(\mathbf{d}|\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\mathbf{u}|, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\tilde{\mathbf{A}}; \theta\right) p\left(\mathbf{b}; \theta\right), \tag{12}$$

where $\theta = [\lambda_{noise}, \lambda_x, \lambda_y, \mu, \nu_x, \nu_y]$ gathers the parameters of the model. Estimation of the model parameters could be obtained through maximization of the marginal distribution of the observations $p\left(\mathbf{d}; \theta\right)$:

$$\hat{\theta} = \arg\max_{\theta} \int\int\int p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) \, d\mathbf{u} \, d\tilde{\mathbf{A}} \, d\mathbf{b}. \tag{13}$$

However, in the present case, this marginalization is not possible, since the posterior of the latent variables given the observations $p(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}|\mathbf{d})$ is not known explicitly and inference via the Expectation-Maximization (EM) algorithm may not be obtained. Thus, we resort to the variational methodology

[4], [7] where we have to maximize a lower bound of $p\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right)$ by employing the *mean field approximation* [7]. The details of the derivation are given in the Appendix. Here we just provide the update equations for the model variables and parameters.

Therefore, in the variational E-step (VE-step) of the algorithm the motion vectors are estimated by:

$$\mathbf{u}_x^{(t+1)} = \lambda_{noise}^{(t)} \mathbf{R}_x^{(t)} \mathbf{B}^{(t)} \mathbf{G}_x \left(\mathbf{d} - \mathbf{G}_y \mathbf{u}_y^{(t)}\right), \tag{14}$$

and

$$\mathbf{u}_y^{(t+1)} = \lambda_{noise}^{(t)} \mathbf{R}_y^{(t)} \mathbf{B}^{(t)} \mathbf{G}_y \left(\mathbf{d} - \mathbf{G}_x \mathbf{u}_x^{(t)}\right), \tag{15}$$

where

$$\mathbf{R}_x^{(t+1)} = \left(\lambda_{noise}^{(t)} \mathbf{G}_x^T \mathbf{B}^{(t)} \mathbf{G}_x + \lambda_x^{(t)} \mathbf{Q}^T \mathbf{A}_x^{(t)} \mathbf{Q}\right)^{-1}, \tag{16}$$

and

$$\mathbf{R}_y^{(t+1)} = \left(\lambda_{noise}^{(t)} \mathbf{G}_y^T \mathbf{B}^{(t)} \mathbf{G}_y + \lambda_y^{(t)} \mathbf{Q}^T \mathbf{A}_y^{(t)} \mathbf{Q}\right)^{-1}. \tag{17}$$

The expectations of the hidden random variables $\boldsymbol{\alpha}_x(i)$ and $\boldsymbol{\alpha}_y(i)$ are updated by:

$$\langle \boldsymbol{\alpha}_k(i) \rangle = \frac{\nu_k^{(t)} + 1}{\nu_k^{(t)} + \lambda_k^{(t)} \left(\left(\mathbf{Q}\mathbf{u}_k^{(t)}\right)_i^2 + \mathbf{C}_k^{(t)}(i,i) \right)}, \tag{18}$$

where $k \in \{x, y\}$, $\left(\mathbf{Q}\mathbf{u}_k^{(t)}\right)_i$ is the i-th element of vector $\left(\mathbf{Q}\mathbf{u}_k^{(t)}\right)$ and the $N \times N$ matrix

$$\mathbf{C}_k^{(t)} = \mathbf{Q}\mathbf{R}_k^{(t)}\mathbf{Q}^T. \tag{19}$$

Notice that $\boldsymbol{\alpha}_k(i)$ is the equivalent parameter present in many variational methods [14], [10] which weights the importance between the data and smoothness term and is generally determined by the user. Here, not only it is updated using the image information but also is spatially varying and has edge-preserving properties by handling abrupt motion changes.

In a similar manner, the expectation of $\mathbf{b}(i)$ is computed by:

$$\langle \mathbf{b}(i) \rangle = \frac{\mu^{(t)} + 1}{\mu^{(t)} + \lambda_{noise}^{(t)} \left(\left(\mathbf{G}\mathbf{u}^{(t)} - \mathbf{d}\right)_i^2 + \mathbf{F}^{(t)}(i,i) \right)}, \tag{20}$$

where $\left(\mathbf{G}\mathbf{u}^{(t)} - \mathbf{d}\right)_i$ is the i-th element of vector $\left(\mathbf{G}\mathbf{u}^{(t)} - \mathbf{d}\right)$ and the $N \times N$ matrix

$$\mathbf{F}^{(t)} = \mathbf{G}_x \mathbf{R}_x^{(t)} \mathbf{G}_x^T + \mathbf{G}_y \mathbf{R}_y^{(t)} \mathbf{G}_y^T. \tag{21}$$

Recall that $\mathbf{b}(i)$ models the residual of the linearization of the brightness constancy constraint using Taylor series expansion and it is updated only from the data.

In (18) and (20) we have omitted the time step index $(t+1)$ from the expectations only for presentation purposes (notation would become *barroque*). The size of matrices $\mathbf{R}_x$, $\mathbf{R}_y$ and consequently $\mathbf{C}_x$, $\mathbf{C}_y$ and $\mathbf{F}$ makes their direct calculation prohibitive. In order to overcome this difficulty, we employ the iterative Lanczos method [22] for their calculation. For matrices $\mathbf{C}_x$, $\mathbf{C}_y$ and $\mathbf{F}$ only the diagonal elements are needed in (18) and (20) and they are obtained as a byproduct of the Lanczos method.

Let us notice that as we can see from (14) and (15), there is a dependency between $\mathbf{u}_x^{(t+1)}$ and $\mathbf{u}_y^{(t)}$, as well as between $\mathbf{u}_y^{(t+1)}$ and $\mathbf{u}_y^{(t)}$. This is also the case in the standard Horn-Schunck method. However, in our approach, all of the involved parameters are computed from the two images.

In the variational M-step (VM-step), where the lower bound is maximized with respect to the model parameters, we obtain:

$$\lambda_{noise}^{(t+1)} = \frac{N}{\sum_{i=1}^{N} \langle \mathbf{b}(i) \rangle \left(\left(\mathbf{G}\mathbf{u}^{(t+1)} - \mathbf{d} \right)_i^2 + \mathbf{F}^{(t+1)}(i,i) \right)}, \tag{22}$$

and equivalently for λ_x and λ_y:

$$\lambda_k^{(t+1)} = \frac{N}{\sum_{i=1}^{N} \langle \boldsymbol{\alpha}_k(i) \rangle \left(\left(\mathbf{Q}\mathbf{u}_k^{(t+1)} \right)_i^2 + \mathbf{C}_k^{(t+1)}(i,i) \right)}, \tag{23}$$

with $k \in \{x, y\}$.

The *degrees of freedom* parameters ν_k of the Student's-*t* distributions are also computed accordingly through the roots of the following equation:

$$\begin{aligned} &\frac{1}{N} \left(\sum_{i=1}^{N} \log \langle \boldsymbol{\alpha}_k(i) \rangle - \sum_{i=1}^{N} \langle \boldsymbol{\alpha}_k(i) \rangle \right) + \digamma \left(\frac{\nu_k^{(t)}}{2} + \frac{1}{2} \right) \\ &- \log \left(\frac{\nu_k^{(t)}}{2} + \frac{1}{2} \right) - \digamma \left(\frac{\nu_k}{2} \right) + \log \left(\frac{\nu_k}{2} \right) + 1 = 0, \end{aligned} \tag{24}$$

for ν_k, $k \in \{x, y\}$, where $\digamma(x)$ is the digamma function (derivative of the logarithm of the Gamma function) and $\nu_k^{(t)}$ is the value of ν_k at the previous iteration.

Finally, by the same procedure we obtain estimates for the parameter μ of the noise distribution

$$\begin{aligned} &\frac{1}{N} \left(\sum_{i=1}^{N} \log \langle \mathbf{b}(i) \rangle - \sum_{i=1}^{N} \langle \mathbf{b}(i) \rangle \right) + \digamma \left(\frac{\mu^{(t)}}{2} + \frac{1}{2} \right) \\ &- \log \left(\frac{\mu^{(t)}}{2} + \frac{1}{2} \right) - \digamma \left(\frac{\mu}{2} \right) + \log \left(\frac{\mu}{2} \right) + 1 = 0. \end{aligned} \tag{25}$$

In our implementation equations (24) and (25) are solved by the bisection method, as also proposed in [16]. The overall algorithm is summarized in Algorithm 1 where initialization of the motion vectors may be obtained by any standard optical flow method. Here we have chosen to use the standard Horn-Schunck algorithm [14].

V. Experimental results

The method proposed herein is a principled Bayesian generalization of the Horn-Schunck (HS) method [14]. Therefore, our purpose is to examine its appropriateness to replace it in the initialization

Algorithm 1 Variational-Bayes optical flow computation

- Initialize $\mathbf{u}_x$, $\mathbf{u}_y$ by the Horn-Schunck optical flow and set $\langle \boldsymbol{\alpha}_x(i) \rangle$, $\langle \boldsymbol{\alpha}_y(i) \rangle$ to the stationary values and $\langle \mathbf{b}(i) \rangle = 0$. Compute $\mathbf{R}_x$, $\mathbf{R}_y$, ν_x, ν_y, μ, λ_x, λ_y, λ_{noise} as solutions to the stationary problem.
- **DO** until convergence
  - **VE-step:**
    * Compute the expectations $\langle \boldsymbol{\alpha}_x(i) \rangle$ and $\langle \boldsymbol{\alpha}_y(i) \rangle$ using (18).
    * Compute the expectation $\langle \mathbf{b}(i) \rangle$ using (20).
  - **VM-step:**
    * Compute λ_{noise} using (22).
    * Compute λ_x, λ_y using (23).
    * Solve for ν_x, ν_y equation (24), using the bisection method.
    * Solve for μ equation (25), using the bisection method.
    * Update matrices $\mathbf{R}_x$ and $\mathbf{R}_y$ in (16) and (17) and the diagonal elements of $\mathbf{C}_x$, $\mathbf{C}_y$ and $\mathbf{F}$ using the Lanczos method.
    * Compute $\mathbf{u}_x$, $\mathbf{u}_y$ from (14) and (15).

of more advanced optical flow schemes. We have also included the well-known and established rival algorithm of Lucas-Kanade (LK) [17]. These are the two methods widely used for initializing more sophisticated optical flow algorithms. Moreover, we have included in the comparison the algorithm proposed in [6], which combines the above two algorithms for feature tracking, based on a framework proposed in [11]. We call this method *Joint Lucas-Kanade* (JLK). To visualize the motion vectors we adopt the color coding figure 2.

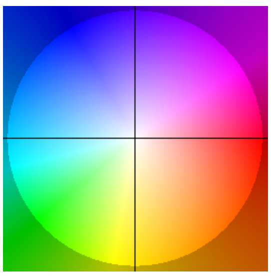

Fig. 2. The optical flow field color-coding. Smaller vectors are lighter and color represents the direction.

The proposed method was tested on image sequences including both synthetic and real scenes. A synthetic sequence included in our experiments consists of two textured triangles moving to different directions (fig. 3(a)). We have synthesized two versions of the sequence: one with equal (*Triangles-equal*) and one with different (*Triangles-unequal*) velocity magnitudes for the triangles in each sequence (the angles of the velocity differ by 90^o in both cases). We have also applied our method to the *Yosemite* sequence (fig. 3(b)) as well as to the *Dimetrodon* sequence (fig. 3(c)) obtained from the Middlebury database [2].

To evaluate the performance of the method two performance measures were computed. The average angular error (AAE) [3] which is the most common measure of performance for optical flow and the average magnitude of difference error (AME) [18]. The latter measure normalizes the errors with respect to the ground truth and ignores normalized error vector norms smaller than a threshold T. We have employed $T = 0.35$ in our evaluation.

The numerical results are summarized in table I, where it may be observed that the method proposed in this paper provides better accuracy with regard to the other methods. More specifically, our algorithm largely outperforms the Lucas-Kanade method and is clearly better than the Horn-Schunck algorithm. Notice that the JLK algorithm is not very accurate as its behavior depends partially on a Lucas-Kanade scheme which fails in all cases (first table row). We conclude that JLK which combines the two approaches may perform better for sparse optical flow applied to features [6] but not for dense flow estimation.

Representative results are presented in figure 3. As it may be seen, our variational-Bayes algorithm provides smooth estimates and simultaneously preserves edge information in the flow field. The Horn-Schunck algorithm has a unique, user determined parameter for controlling the relative weight of data and smoothness terms and cannot be as accurate as the newly proposed approach. Moreover, this parameter is not spatially varying, thus providing results of lower quality (see for instance the *Dimetrodon* sequence results in fig. 3(c)). Finally, notice the dilated motion field of the Lucas-Kanade algorithm in fig. 3(a).

TABLE I

OPTICAL FLOW ERRORS FOR THE COMPARED METHODS.

Method	*Triangles-Equal*		*Triangles-Unequal*		*Yosemite*		*Dimetrodon*	
	AAE	AME	AAE	AME	AAE	AME	AAE	AME
Lucas-Kanade [17]	5.91°	0.15	8.58°	0.17	11.65°	0.26	27.52°	0.56
Horn-Schunck [14]	2.47°	0.05	5.57°	0.14	5.43°	0.13	8.50°	0.49
JLK [6]	4.10°	0.07	6.95°	0.18	7.97°	0.18	33.14°	0.65
Proposed method	**1.06°**	**0.02**	**3.93°**	**0.10**	**4.45°**	**0.12**	**4.31°**	**0.13**

Furthermore, the above comments are also confirmed by the cumulative histograms for the AAE

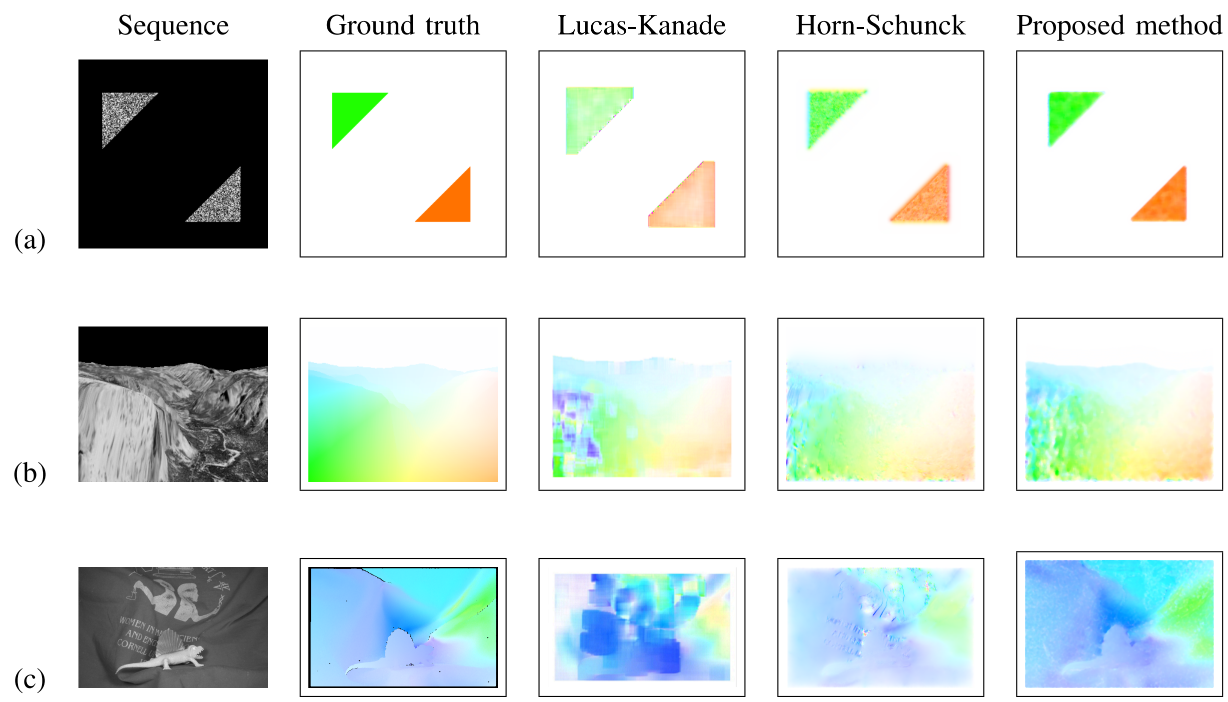

Fig. 3. Representative optical flow results following the coding in fig. 2 for the sequences: (a) *Triangles-equal* , (b) *Yosemite* and (c) *Dimetrodon*.

and AME for all of the compared algorithms, shown in figure 4. A point on the curve represents the percentage of optical flow errors that are less or equal than the value of the respective error on the horizontal axis. The higher the curve the better is the performance of the method. An ideal performance would provide a curve parallel to the horizontal axis, meaning that all of the errors are zero.

The proposed method is a Bayesian generalization of the Horn-Schunck algorithm [14] and therefore it carries the limits and drawbacks of the original approach. The contribution of the method proposed here is to substitute the standard HS algorithm by its variational version in other optical flow methods where the mother algorithm is used as an initialization step. The experiments showed that this is worth performing as there ia a clear gain in accuracy.

The algorithm takes on average less than a minute to converge o a standard PC running MATLAB, depending on the number of image pixels (e.g. it takes 80 seconds for the 584×388 sized *Dimetrodon* sequence). More than half of this time is due to the Lanczos method used for diagonalizing the matrices in eq. (16)-(17).

VI. Conclusion

The optical flow estimation method proposed in this paper relies on a probabilistic formulation of the problem along with a variational Bayesian inference approach. The spatially varying Student's

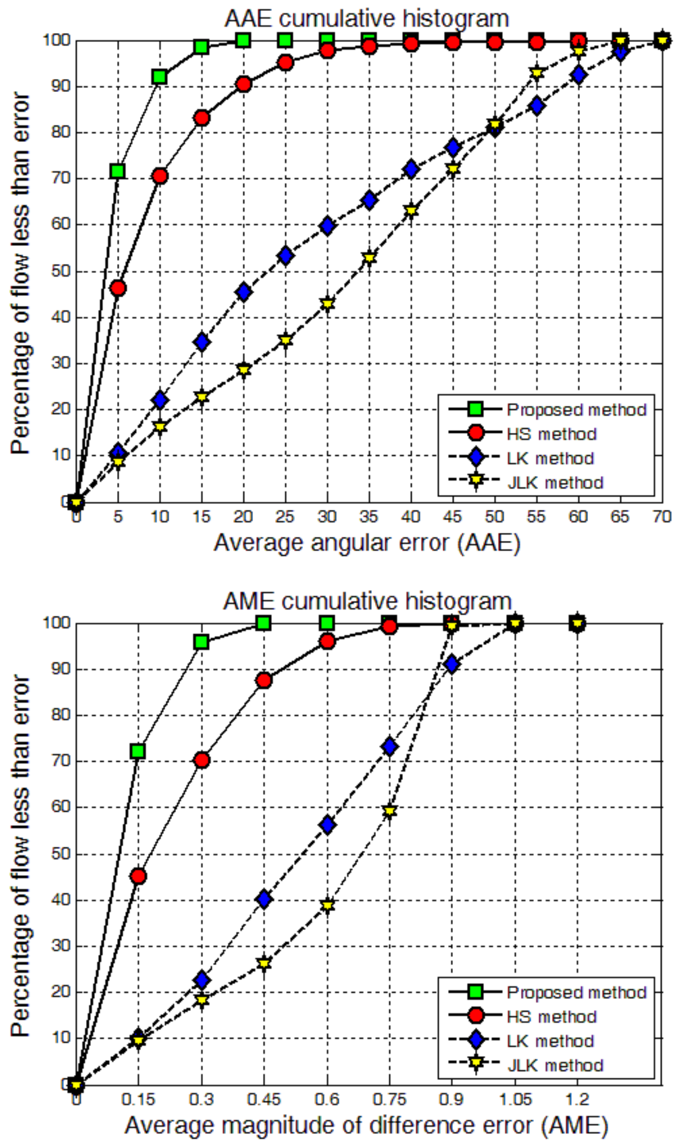

Fig. 4. Performances of the compared algorithms on the *Dimetrodon* sequence [2]. Cumulative histograms showing the percentage of the optical flow errors which are lower than a given value (represented along the horizontal axis) for the AAE (top) and the AME (bottom).

t-distribution of the motion vectors achieves selective application of smoothness leaving motion edges unaffected. Furthermore, any residuals of the linearization of the brightness constancy constraint are also modeled leading to better accuracy.

A perspective of this study is to extend the variational-Bayes framework to other standard variational methods incorporating more sophisticated constraints on the motion field, like the method in [10] or the algorithm proposed in [11]. We believe that both of these methods could benefit from the fully Bayesian framework.

APPENDIX A

In what follows we present in detail the derivation of the update equations for the model variables and parameters.

In the fully Bayesian framework, the complete data likelihood, including the hidden variables and the parameters of the model, is given by:

$$p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) = p\left(\mathbf{d}|\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\mathbf{u}|, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\tilde{\mathbf{A}}; \theta\right) p\left(\mathbf{b}; \theta\right), \tag{26}$$

where $\theta = [\lambda_{noise}, \lambda_x, \lambda_y, \mu, \nu_x, \nu_y]$ gathers the parameters of the model. Estimation of the model parameters could be obtained through maximization of the marginal distribution of the observations $p\left(\mathbf{d}; \theta\right)$:

$$\hat{\theta} = \arg\max_{\theta} \int\int\int p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) d\mathbf{u} d\tilde{\mathbf{A}} d\mathbf{b}. \tag{27}$$

However, in the present case, this marginalization is not possible, since the posterior of the latent variables given the observations $p(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}|\mathbf{d})$ is not known explicitly and inference via the Expectation-Maximization (EM) algorithm may not be obtained. Thus, we resort to the variational methodology [4], [7] where we have to maximize a lower bound of $p\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right)$:

$$L\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) = \int_{\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}} q\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right) \log \frac{p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right)}{q\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right)}. \tag{28}$$

This involves finding approximations of the posterior distribution of the hidden variables, denoted by $q\left(\mathbf{u}\right)$, $q\left(\tilde{\mathbf{A}}\right)$, $q\left(\mathbf{b}\right)$ because there is no analytical form of the auxiliary function q for which the bound in (28) becomes equality. In the variational methodology, however, we employ the *mean field* approximation [7]:

$$q\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right) = q\left(\mathbf{u}\right) q\left(\tilde{\mathbf{A}}\right) q\left(\mathbf{b}\right), \tag{29}$$

and (28) becomes:

$$L\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) = \int_{\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}} q\left(\mathbf{u}\right) q\left(\tilde{\mathbf{A}}\right) q\left(\mathbf{b}\right) \log \frac{p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right)}{q\left(\mathbf{u}\right) q\left(\tilde{\mathbf{A}}\right) q\left(\mathbf{b}\right)} \tag{30}$$

In our case, in the E-step of the variational algorithm (VE-step), optimization of the functional $L\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right)$ is performed with respect to the auxiliary functions. Following the variational inference framework, the distributions $q\left(\mathbf{u}_k\right)$, $k \in x,\ y$, are Normal:

$$q\left(\mathbf{u}\right) = \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_x \\ \mathbf{m}_y \end{bmatrix}, \begin{bmatrix} \mathbf{R}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_y \end{bmatrix}\right), \tag{31}$$

yielding

$$q\left(\mathbf{u}_x\right) = \mathcal{N}\left(\mathbf{m}_x, \mathbf{R}_x\right), \tag{32}$$

and

$$q\left(\mathbf{u}_y\right)=\mathcal{N}\left(\mathbf{m}_y,\mathbf{R}_y\right). \tag{33}$$

Therefore, this bound is actually a function of the parameters $\mathbf{R}_k$ and $\mathbf{m}_k$, $k\in\{x,y\}$ and a functional with respect to the auxiliary functions $q(\mathbf{a}_k), q(\mathbf{b})$. Using (29), the variational bound in our problem becomes:

$$\begin{aligned}
&L\left(q(\mathbf{u}_x),q(\mathbf{u}_y),q(\mathbf{a}_x),q(\mathbf{a}_y),q(\mathbf{b}),\theta_1,\theta_2\right)=\\
&\int\!\!\int\!\!\int\left(\prod_{k\in\{x,y\}}q(\mathbf{u}_k;\theta_1)q(\mathbf{a}_k)\right)q(\mathbf{b})\log p(\mathbf{d},\mathbf{u},\tilde{\mathbf{A}},\mathbf{b};\theta_2)d\mathbf{u}\ d\tilde{\mathbf{A}}d\mathbf{b}\\
&-\int\!\!\int\!\!\int\left(\prod_{k\in\{x,y\}}q(\mathbf{u}_k;\theta_1)q(\mathbf{a}_k)\right)q(\mathbf{b})\log\left(\left(\prod_{k\in\{x,y\}}p(\mathbf{u}_k;\theta_1)q(\mathbf{a}_k)\right)q(\mathbf{b})\right)d\mathbf{u}\ d\tilde{\mathbf{A}}d\mathbf{b}
\end{aligned} \tag{34}$$

where we have separated the parameters into two sets: $\theta_1=[\mathbf{R}_x,\mathbf{R}_x,\mathbf{m}_x,\mathbf{m}_y]$ and $\theta_2=[\mathbf{a}_x,\mathbf{a}_y,\mathbf{b},\lambda_x,\lambda_y,\nu_x,\nu_y]$. Thus, in the VE-step of our algorithm the bound must be optimized with respect to $\mathbf{R}_k$, $\mathbf{m}_k$, $q(\mathbf{a}_k)$ and $q(\mathbf{b})$.

Taking the derivative of (34) with respect to $\mathbf{m}_k$, $\mathbf{R}_k$, $q(\boldsymbol{\alpha}_k)$ and $q(\mathbf{b})$ and setting the result equal to zero, we obtain the following update equations:

$$\mathbf{m}_x^{(t+1)}=\lambda_{noise}^{(t)}\mathbf{R}_x^{(t)}\hat{\mathbf{B}}^{(t)}\mathbf{G}_x\left(\mathbf{d}-\mathbf{G}_y\mathbf{u}_y^{(t)}\right), \tag{35}$$

and

$$\mathbf{m}_y^{(t+1)}=\lambda_{noise}^{(t)}\mathbf{R}_y^{(t)}\hat{\mathbf{B}}^{(t)}\mathbf{G}_y\left(\mathbf{d}-\mathbf{G}_x\mathbf{u}_x^{(t)}\right), \tag{36}$$

where

$$\mathbf{R}_x^{(t+1)}=\left(\lambda_{noise}^{(t)}\mathbf{G}_x^T\hat{\mathbf{B}}^{(t)}\mathbf{G}_x+\lambda_x^{(t)}\mathbf{Q}^T\hat{\mathbf{A}}_x^{(t)}\mathbf{Q}\right)^{-1}, \tag{37}$$

and

$$\mathbf{R}_y^{(t+1)}=\left(\lambda_{noise}^{(t)}\mathbf{G}_y^T\hat{\mathbf{B}}^{(t)}\mathbf{G}_y+\lambda_y^{(t)}\mathbf{Q}^T\hat{\mathbf{A}}_y^{(t)}\mathbf{Q}\right)^{-1}, \tag{38}$$

Notice that the final estimates for $\mathbf{u}_x$, $\mathbf{u}_y$ are $\mathbf{m}_x$ and $\mathbf{m}_y$, in (14) and (15), respectively.

After some manipulation, we obtain the update equations for the model parameters which maximize (34) with respect to $q\left(\mathbf{a}_k\right)$, $q\left(\mathbf{b}\right)$. The form of all q approximating-to-the-posterior functions will remain the same as the corresponding prior (due to the conjugate priors we employ) namely $q\left(\mathbf{a}_k\right)$, $q\left(\mathbf{b}\right)$ which approximate $p\left(\mathbf{a}_k|\mathbf{u}_k,\lambda_k,\mathbf{C}_k;\nu_k\right)$, $p(\mathbf{b}|\mathbf{u},\lambda_{noise},\mathbf{F};\mu)$ will follow Gamma distributions, $\forall i=1,...,N,\forall k\in\{x,y\}$:

$$q^{(t+1)}(\boldsymbol{\alpha}_k(i))=Gamma\left[\frac{\nu_k^{(t)}}{2}+\frac{1}{2},\frac{\nu_k^{(t)}}{2}+\frac{1}{2}\lambda_k^{(t)}\left(\left(\mathbf{Q}\mathbf{u}_k^{(t)}\right)_i^2+\mathbf{C}_k^{(t)}(i,i)\right)\right], \tag{39}$$

and

$$q^{(t+1)}(\mathbf{b}(i))=Gamma\left[\frac{\mu^{(t)}}{2}+\frac{1}{2},\frac{\mu^{(t)}}{2}+\frac{1}{2}\lambda_{noise}^{(t)}\left(\left(\mathbf{G}\mathbf{u}^{(t)}-\mathbf{d}\right)_i^2+\mathbf{F}^{(t)}(i,i)\right)\right], \tag{40}$$

where the $N \times N$ matrix

$$\mathbf{C}_k^{(t)} = \mathbf{Q}\mathbf{R}_k^{(t)}\mathbf{Q}^T, \tag{41}$$

and the $N \times N$ matrix

$$\mathbf{F}^{(t)} = \mathbf{G}_x\mathbf{R}_x^{(t)}\mathbf{G}_x^T + \mathbf{G}_y\mathbf{R}_y^{(t)}\mathbf{G}_y^T. \tag{42}$$

The size of matrices $\mathbf{R}_x$, $\mathbf{R}_y$ and consequently $\mathbf{C}_x$, $\mathbf{C}_y$ and $\mathbf{F}$ makes their direct calculation prohibitive. In order to overcome this difficulty, we employ the iterative Lanczos method [22] for their calculation. For matrices $\mathbf{C}_x$, $\mathbf{C}_y$ and $\mathbf{F}$ only the diagonal elements are needed in (39) and (40) and they are obtained as a byproduct of the Lanczos method.

Let us notice that as we can see from (14) and (15), there is a dependency between $\mathbf{u}_x$ and $\mathbf{u}_y$, as it is the case in the standard Horn-Schunck method.

Notice also that since each $q^{(t+1)}\left(\boldsymbol{\alpha}_k(i)\right)$ is a Gamma pdf, it is easy to derive its expected value:

$$\langle \boldsymbol{\alpha}_k(i) \rangle_{q^{(t+1)}(\boldsymbol{\alpha}_k(i))} = \frac{\nu_k^{(t)} + 1}{\nu_k^{(t)} + \lambda_k^{(t)}\left(\left(\mathbf{Q}\mathbf{u}_k^{(t)}\right)_i^2 + \mathbf{C}_k^{(t)}(i,i)\right)}, \tag{43}$$

and the same stands for the expected value of $\mathbf{b}(i)$:

$$\langle \mathbf{b}(i) \rangle_{q^{(t+1)}(\mathbf{b}(i))} = \frac{\mu^{(t)} + 1}{\mu^{(t)} + \lambda_{noise}^{(t)}\left(\left(\mathbf{G}\mathbf{u}^{(t)} - \mathbf{d}\right)_i^2 + \mathbf{F}^{(t)}(i,i)\right)}, \tag{44}$$

where $\langle . \rangle_{q(.)}$ denotes the expectation with respect to an arbitrary distribution q(.). These estimates are used in (14), (15), (16) and (17), where $\hat{\mathbf{A}}_k^{(t)}$ and $\hat{\mathbf{B}}^{(t)}$ are diagonal matrices with elements:

$$\hat{\mathbf{A}}_k^{(t)}(i,i) = \langle \boldsymbol{\alpha}_k(i) \rangle_{q^{(t)}(\boldsymbol{\alpha}_k(i))},$$

and

$$\hat{\mathbf{B}}^{(t)}(i,i) = \langle \mathbf{b}(i) \rangle_{q^{(t)}(\mathbf{b}(i))},$$

for $i = 1, \ldots, N$. At the variational M-step, the bound is maximized with respect to the model parameters:

$$\theta_2^{(t+1)} = \arg\max_{\theta_2} L\left(q^{(t+1)}\left(\mathbf{u}_k\right), q^{(t+1)}\left(\hat{\mathbf{A}}_k\right), q^{(t+1)}\left(\hat{\mathbf{B}}\right), \theta_1^{(t+1)}, \theta_2\right), \tag{45}$$

where

$$\begin{aligned} &L\left(q^{(t+1)}\left(\mathbf{u}_k\right), q^{(t+1)}\left(\hat{\mathbf{A}}_k\right), q^{(t+1)}\left(\hat{\mathbf{B}}\right), \theta_1^{(t+1)}, \theta_2\right) \propto \\ &\langle \log p\left(\mathbf{d}, \mathbf{u}, \hat{\mathbf{A}}_k, \hat{\mathbf{B}}; \theta_2\right) \rangle_{q\left(\mathbf{u}_k; \theta_1^{(t+1)}\right), q^{(t+1)}\left(\hat{\mathbf{A}}_k\right), q^{(t+1)}\left(\hat{\mathbf{B}}\right)} \end{aligned} \tag{46}$$

is calculated using the results from (14) - (40).

The update for λ_{noise} is obtained after taking the derivative of

$$L\left(q^{(t+1)}\left(\mathbf{u}_k\right), q^{(t+1)}\left(\hat{\mathbf{A}}_k\right), q^{(t+1)}\left(\hat{\mathbf{B}}\right), \theta_1^{(t+1)}, \theta_2\right)$$

in (34) with respect to it and setting it to zero:

$$\lambda_{noise}^{(t+1)} = \frac{N}{\sum_{i=1}^{N} \mathbf{b}^{(t+1)}(i) \left(\left(\mathbf{G}\mathbf{u}^{(t+1)} - \mathbf{d} \right)_i^2 + \mathbf{F}^{(t+1)}(i,i) \right)}, \tag{47}$$

By the same means we obtain the estimates for λ_x and λ_y:

$$\lambda_k^{(t+1)} = \frac{N}{\sum_{i=1}^{N} \boldsymbol{\alpha}_k^{(t+1)}(i) \left(\left(\mathbf{Q}\mathbf{u}_k^{(t+1)} \right)_i^2 + \mathbf{C}_k^{(t+1)}(i,i) \right)}, \tag{48}$$

with $k \in \{x, y\}$.

The *degrees of freedom* parameters ν_k of the Student's-*t* distributions are also computed accordingly through the roots of the following equation:

$$\begin{aligned} &\frac{1}{N} \left(\sum_{i=1}^{N} \log \langle \boldsymbol{\alpha}_k(i) \rangle_{q^{(t+1)}(\mathbf{A}_k)} - \sum_{i=1}^{N} \langle \boldsymbol{\alpha}_k(i) \rangle_{q^{(t+1)}(\mathbf{A}_k)} \right) + \digamma \left(\frac{\nu_k^{(t)}}{2} + \frac{1}{2} \right) \\ &- \log \left(\frac{\nu_k^{(t)}}{2} + \frac{1}{2} \right) - \digamma \left(\frac{\nu_k}{2} \right) + \log \left(\frac{\nu_k}{2} \right) + 1 = 0, \end{aligned} \tag{49}$$

for ν_k, $k \in \{x, y\}$, where $\digamma(x)$ is the digamma function (derivative of the logarithm of the Gamma function) and $\nu_k^{(t)}$ is the value of ν_k at the previous iteration.

Finally, by the same procedure we obtain estimates for the parameter μ of the noise distribution

$$\begin{aligned} &\frac{1}{N} \left(\sum_{i=1}^{N} \log \langle \mathbf{b}(i) \rangle_{q^{(t+1)}(\mathbf{b}(i))} - \sum_{i=1}^{N} \langle \mathbf{b}(i) \rangle_{q^{(t+1)}(\mathbf{b}(i))} \right) + \digamma \left(\frac{\mu^{(t)}}{2} + \frac{1}{2} \right) \\ &- \log \left(\frac{\mu^{(t)}}{2} + \frac{1}{2} \right) - \digamma \left(\frac{\mu}{2} \right) + \log \left(\frac{\mu}{2} \right) + 1 = 0. \end{aligned} \tag{50}$$

In our implementation equations (49) and (50) are solved by the bisection method, as also proposed in [16].

References

[1] L. Alvarez, J. Weickert, and J. Sanchez. Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39:41–56, 2000.

[2] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *Proceedings of the International Conference of Computer Vision, (ICCV)*, pages 1–8, 2007.

[3] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision, (IJCV)*, 12(1):43–77, 1994.

[4] M. J. Beal. The variational Bayesian EM algorithm for incomplete data with application to scoring graphical model structures. *Bayesian Statistics*, 7:453–464, 2003.

[5] R. Ben-Ari and N. Sochen. A general framework and new alignment criterion for dense optical flow. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:529–536, 2006.

[6] S. Birchfield and S. Pundlik. Joint tracking of features and edges. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[7] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[8] M. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, 1996.

[9] M. Black and D. Fleet. Probabilistic detection and tracking of motion boundaries. *International Journal of Computer Vision*, 38:231–245, 200.

[10] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proceedings of the 8th European Conference on Computer Vision, (ECCV)*, volume 4, pages 25–36, 2004.

[11] A. Bruhn, J. Weickert, and C. Schnorr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.

[12] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders. Variational Bayesian image restoration based on a product of t-distributions image prior. *IEEE Transactions on Image Processing*, 17(10):1795–1805, 2008.

[13] B. Glocker, T. H. Heibel, N. Navab, P. Kohli, and C. Rother. Triangleflow: optical flow with triangulation-based higher-order likelihoods. In *Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III*, ECCV'10, pages 272–285, 2010.

[14] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[15] K. J. Lee, D. Kwon, I. D. Yun, and S. U. Lee. Optical flow estimation with adaptive convolution kernel prior on discrete framework. In *CVPR*, pages 2504–2511, 2010.

[16] C. Liu and D. B. Rubin. ML estimation of the t-distribution using EM and its extensions. Technical report, ECM and ECME, Statistica Sinica, 1995.

[17] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intellegence, (IJCAI)*, pages 674–679, 1981.

[18] B. McCane, K. Novins, D. Crannitch, and B. Galvin. On benchmarking optical flow. *Computer Vision and Image Understanding*, 84(1):126–143, 2001.

[19] E. Memin and P. Perez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5):703 –719, 1998.

[20] R. Molina. On the hierarchical Bayesian approach to image restoration. applications to astronomical images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(11):1122–1128, 1994.

[21] H. Nagel and W. Enkelman. An investigation of smoothness constraints for estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(5):565–593, 1986.

[22] C. Paige and M. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis*, 12:617–629, 1975.

[23] X. Ren. Local grouping for optical flow. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[24] S. Roth and M. Black. On the spatial statistics of optical flow. *International Journal of Computer Vision*, 74(1):33–50, 2007.

[25] D. Sun, S. Roth, J. P. Lewis, and M. J. Black. Learning optical flow. In *Proceedings of the 10th European Conference on Computer Vision (ECCV): Part III*, ECCV '08, pages 83–97, 2008.

[26] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure- and motion-adaptive regularization for high accuracy optic flow. In *International Conference on Computer Vision (ICCV)*, pages 1663–1668, 2009.

[27] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, 2001.

[28] J. Weickert and C. Schnörr. Variational optical flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14(3):245–255, 2001.

[29] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *CVPR*, pages 2464–2471, 2010.

[30] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–

388, 2011.

[31] C. Zitnick, N. Jojic, and S. Kang. Consistent segmentation for optical flow estimation. In *Proceedings of the International Conference on Computer Vision, (ICCV)*, volume 2, pages 1308–1315, 2005.