Variational-Bayes Optical Flow

Giannis Chantas, Theodosios Gkamas and Christophoros Nikou*

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, i.e., the prior is a multivariate Student's t-distribution, while the only observations available is the temporal and spatial 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,

G. Chantas is with the Information Technologies Institute, Centre of Research and Technology. HELLAS, 57001 Thessaloniki, Greece, (email: gchantas@iti.gr).

T. Gkamas is with the Laboratoire des Sciences de l'Ingenieur, de l'Informatique et de l'Imagerie (iCube), CNRS UMR 7357, University of Strasbourg, France, (email: tgkamas@gmail.com).

C. Nikou is with the Department of Computer Science and Engineering, University of Ioannina, 45110 Ioannina, Greece, (email: cnikou@cs.uoi.gr, webpage: www.cs.uoi.gr/~cnikou).

Asterisk indicates corresponding author

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 [24]. 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 [18], 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 that treat outliers in both matching and smoothness assumptions [9], [10], [26], variational methodologies [1], [6], [22] that incorporate temporal smoothness constraints and gradient constancy assumptions [12], [40], the integration of spatial priors [31], the segmentation of the image pixels or the motion vectors [30], [41], [45], and learning from ground truth data [36]. Moreover, efforts to combine local and global adaptive techniques were also proposed [16], [38], such as the technique in [13], where the motion vectors are smoothed before being forwarded to a global scheme or the method in [7], where the estimated motion of a feature is influenced by the estimated motion of its neighbors. In both of these methods [13], [7] the spatial integration is isotropic while an anisotropic smoothness term which works complementary with the data term was also conceived [44].

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 [12], this issue should be thoroughly considered.

The smoothness term assumes global or piecewise smoothness spatially. Its properties may vary from homogeneous and isotropic [18], to inhomogeneous [12], or even simultaneously inhomogeneous and anisotropic [28], [39].

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

2

the user remaining fixed during the whole process. This is the case not only for the early algorithm of Horn-Schunck [18] but also for the latest versions of this category of methods [7], [12], [13]. 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 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 [8], [37] which yields a variational expectation-maximization (EM) framework. The variational EM algorithm has revealed to be successful in image analysis applications of mixture models [37], such as natural image segmentation [33], [34], [35] and poliarimetric image restoration [32].

Many works are similar with the philosophy of the proposed work. Next, we mention the most important of them. The work in [43] is similar with our work in the sense that the Bayesian methodology is utilized. However, in [43], the approximation used to overcome intractability proceeds as follows. The independency assumption in the prior, which is necessary for a tractable estimation scheme, applies on the pixel level, which means that different pixels neighborhoods are treated as independent random variables. In our work, however, the independency assumption is different: u and b are independent in the inferred posterior.

In [17] a hybrid maximum a posteriori estimation (MAP) and variational Bayesian formulation is proposed for the optical flow problem, while, in our work, the MAP approach is not used at all. This means that, in our work, posteriors for u and b are inferred and estimation of the motion field is

based on the posteriors of the rest model random variables (and not only on their point estimates), while in [17] the posterior of a set of random variables does not play a role for the estimation of the other. Instead, only point estimates are used for this purpose. Moreover, key role in [17] Bayesian model selection is employed. The model selection is performed in a Bayesian framework, where the Gaussian assumption is adopted for the sake of tractable computations. However, in our work, model selection does not take place and the form of the prior and data distributions are assumed known. This makes our problem less difficult, without compromising the efficiency of the optical flow estimation.

The works in [20], [19], [21] are similar with the proposed work, in the sense that robust distributions for prior and observation models are used. More specifically, the data likelihood and the prior are heavy tailed distributions, providing robust estimation. The intractable problem of parameter estimation, however, is bypassed by using the Gaussian approximation (using Taylor expansion), in contrast with our variational (mean field approximation) methodology. Furthermore, in our work, estimation robustness is achieved by the spatially adaptive nature of the distribution adopted for modeling the random variables. Robust motion field estimation is also used in [42] using the total-variation criterion.

Lastly, the model adopted for the optical flow problem in [11] is very similar with the proposed model in this work. More specifically, the prior used for the motion field is almost identical with that proposed herein. However, the main difference is that in [11] a MAP approach is followed, in contrast with the variational Bayesian methodology we adotp in this work. The advantages of this methodology are presented in [37]. The main difference with MAP in our case is that the hidden variables covariance is also estimated, in addition to their mean, and this influences the inference process. Lastly, the model parameters are also estimated in this principled setting, which is very crucial for the algorithm automatic functioning.

By setting the optical flow problem up in a variational 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 [18], 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 [24], [18] which are used to initialize more sophisticated methods, but also to a recently proposed version of their joint combination [7].

A preliminary short version of this work was presented in [15], where the main idea was highlighted without showing the variational EM framework. Here, we describe the model in full detail, with the derivation of the update equations for all of the model hidden variables and parameters along with

more experimental results.

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. In order to define the prior for $\mathbf{u}_k, k = 1, 2$, we denote by $\epsilon_k, \ \forall k \in \{x, y\}$ the $N \times 1$ vectors:

$$\epsilon_k = \mathbf{Q}\mathbf{u_k}, \ \forall k \in \{x, y\}, \tag{2}$$

where $\mathbf{Q}$ is the $N \times N$ Laplacian operator matrix. We adopt the assumption that each $\epsilon_k(i)$ follows a Student's-*t* distribution:

$$\epsilon_k(i) = \sim \mathcal{S}t\left(0, \lambda_k, \nu_k\right), \forall i = 1, ..., N, \forall k \in \{x, y\}. \tag{3}$$

The Student's-*t* distribution implies a two-level generative process [8]. More specifically, at the first level, the random variables $\alpha_k(i), k \in \{x, y\}$ are drawn from two independent Gamma distributions:

$$\alpha_k(i) \sim \text{Gamma}\left(\frac{\nu_k}{2}, \frac{\nu_k}{2}\right) \tag{4}$$

At the second level, given $\alpha_k(i)$, $\epsilon_k(i)$ is produced by a zero mean Gaussian distribution, with variance $\lambda_k \alpha_k(i)$:

$$p(\epsilon_k(i)|\alpha_k(i)) = \mathcal{N}\left(0, (\lambda_k \alpha_k(i))^{-1}\right). \tag{5}$$

This is the conditional distribution of $\epsilon_k(i)$.

5

Equation (3), based on (4) and (3), gives the marginal distribution $p(\epsilon_k(i))$, which is Student's t, that can be written as the integral:

$$p(\epsilon_k(i)) = \int_0^\infty p(\epsilon_k|\alpha_k(i)) \, p(\alpha_k(i)) \, d\alpha_k(i). \tag{6}$$

As we will see next, $p(\epsilon_k(i))$ implies also a Student's-t distribution for the motion field. Thus, it is desirable to analyze this distribution when used as prior. The variables $\alpha_k(i)$ are hidden because they are not apparent in (6) since they have been integrated out. As the *degrees of freedom* parameter $\nu_k \to \infty$, the pdf of $\alpha_k(i)$ 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 $\alpha_k(i) = 1$ 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". 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 [27], [14]. However, using this type of modeling, we can avoid smoothing out abrupt motion fields, based on the robustness property of the Student's-t distribution.

By defining the $N \times N$ diagonal matrices

$$\mathbf{A}_k = \mathrm{diag}[\alpha_k(1), \dots, \alpha_k(N)]^T, k \in \{x, y\}, \tag{7}$$

and substituting ϵ_k with $\mathbf{Q}\mathbf{u}_k$ (according to (2)) in (5), the conditional pdf of the horizontal and vertical motion fields may now be expressed by:

$$p(\mathbf{u}_k|\mathbf{A}_k) = \mathcal{N}\left(\underline{\mathbf{0}}, \left(\lambda_k \mathbf{Q}^T \mathbf{A}_k \mathbf{Q}\right)^{-1}\right), \tag{8}$$

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 conditional (given $\mathbf{A}$) pdf of the motion field $\mathbf{u} = [\mathbf{u}_x, \mathbf{u}_y]^T$ is given by:

$$p(\mathbf{u}|\mathbf{A}) = p(\mathbf{u}_x|\mathbf{A}_x) p(\mathbf{u}_y|\mathbf{A}_y), \tag{9}$$

or equivalently:

$$p\left(\mathbf{u}|\tilde{\mathbf{A}}\right) = \mathcal{N}\left([\underline{\mathbf{0}}, \underline{\mathbf{0}}]^T \left(\tilde{\mathbf{Q}}^T \tilde{\mathbf{A}} \tilde{\mathbf{Q}}\right)^{-1}\right), \tag{10}$$

$\tilde{\mathbf{A}}$ is a $2N \times 2N$ matrix

$$\tilde{\mathbf{A}} = \begin{bmatrix} \lambda_x \mathbf{A}_x & \mathbf{0} \\ \mathbf{0} & \lambda_y \mathbf{A}_y \end{bmatrix}, \tag{11}$$

$\tilde{\mathbf{Q}}$ is a $2N \times 2N$ matrix

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} \tag{12}$$

and $\mathbf{0}$ is a zero matrix of size $N \times N$.

Notice that the Gaussian distribution in (10) is zero-mean. This choice is arbitrary, since it plays no role in the modeling and inference procedure. This stems from the fact that whatever the increment of the motion field $\mathbf{u}$ the Laplacian operator $\mathbf{Q}$ cancels it, since it holds:

$$\mathbf{Q}(\mathbf{u}_k + \mathbf{c}) = \mathbf{Q}\mathbf{u} = \boldsymbol{\epsilon}_k,$$

where $\mathbf{c}$ is an $N \times 1$ vector with all its elements equal to a constant number, i.e., the increment.

III. A PROBABILISTIC MODEL FOR OPTICAL FLOW

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

$$\mathbf{G}\mathbf{u} = \mathbf{d}, \tag{13}$$

where the block diagonal $N \times 2N$ matrix

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_x & \mathbf{G}_y \end{bmatrix}, \tag{14}$$

with

$$\mathbf{G}_x = \text{diag} \left[\frac{\partial \mathbf{I}(\mathbf{x}_1)}{\partial x}, \ldots, \frac{\partial \mathbf{I}(\mathbf{x}_N)}{\partial x} \right]^T, \tag{15}$$

$$\mathbf{G}_y = \text{diag} \left[\frac{\partial \mathbf{I}(\mathbf{x}_1)}{\partial y}, \ldots, \frac{\partial \mathbf{I}(\mathbf{x}_N)}{\partial y} \right]^T, \tag{16}$$

contains the spatial derivatives in the horizontal and vertical directions lexicographically and the $N \times 1$ vector

$$\mathbf{d} = [\mathbf{I}(\mathbf{x}_1) - \mathbf{J}(\mathbf{x}_1), \ldots, \mathbf{I}(\mathbf{x}_N) - \mathbf{J}(\mathbf{x}_N)]^T, \tag{17}$$

contains the temporal image differences. Therefore, to visually highlight the role of matrix $\mathbf{G}$, eq. (13) may be also written as:

$$\begin{bmatrix} \mathbf{G}_x & \mathbf{G}_y \end{bmatrix} \begin{bmatrix} \mathbf{u}_x \\ \mathbf{u}_y \end{bmatrix} = \mathbf{d}. \tag{18}$$

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 (13) yielding:

$$\mathbf{G}\mathbf{u} + \mathbf{w} = \mathbf{d}. \tag{19}$$

We also assume spatially varying Student's-t statistics for this $N \times 1$ noise vector. Thus, similarly with the case of (u), the conditional of $\mathbf{w}$ is a Gaussian:

$$\mathbf{w} \sim \mathcal{N}\left(\underline{\mathbf{0}}, (\lambda_{noise}\mathbf{B})^{-1}\right), \tag{20}$$

where $\lambda_{noise}\mathbf{B}$ is the noise precision matrix and

$$\mathbf{B} = \text{diag}[\mathbf{b}(1), \ldots, \mathbf{b}(N)]^T. \tag{21}$$
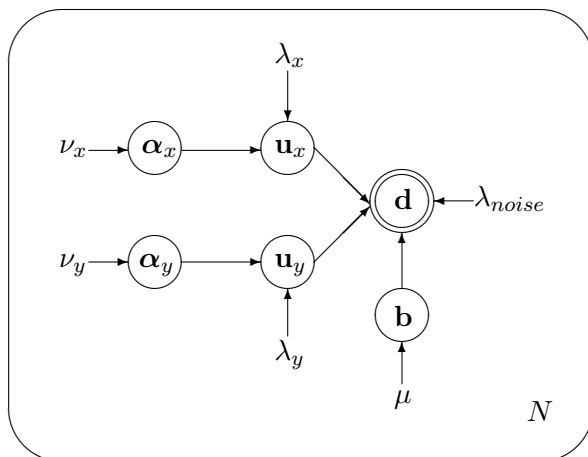
Fig. 1. Graphical model for the optical flow problem. The hidden variables of the model are the optical flow unknowns $\mathbf{u}_x$ and $\mathbf{u}_y$, their precisions $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ and the residual of the Taylor series approximation of the optical flow equation $\mathbf{b}$. The only observation is vector $\mathbf{d}$, containing the temporal image differences, which is denoted by the double circle. Finally, λ_x, λ_y, λ_{noise}, ν_x, ν_y and μ are the model's parameters. All of the variables and the observations are of dimension N. However, the problem is ill-posed as there are $2N$ unknowns (the elements of $\mathbf{u}_x$ and $\mathbf{u}_y$) and only N observations (the elements of $\mathbf{d}$).

Also, the *t*-distribution implies that each $\mathbf{b}(i)$, $i = 1, .., N$ is Gamma distributed with parameter μ:

$$\mathbf{b}(i) \sim \text{Gamma}\left(\frac{\mu}{2}, \frac{\mu}{2}\right). \tag{22}$$

Following the optical flow matrix-vector formulation in (19) and the noise modeling in (20) and (22), we arrive at the probability of the temporal image differences given the motion vectors:

$$p(\mathbf{d}|\mathbf{u}) = \mathcal{N}\left(\mathbf{Gu}, (\lambda_{noise}\mathbf{B})^{-1}\right). \tag{23}$$

The above probabilistic formulation of the optical flow problem is represented by the graphical model of fig. 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

$$p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) = p\left(\mathbf{d}|\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\mathbf{u}|\tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\tilde{\mathbf{A}}; \theta\right) p\left(\mathbf{b}; \theta\right), \tag{24}$$

where

$$\theta = \{\lambda_{noise}, \lambda_x, \lambda_y, \mu, \nu_x, \nu_y\}, \tag{25}$$

gathers the parameters of the model. Estimation of the model parameters could be obtained through maximization of the marginal distribution of the observations $p(\mathbf{d}; \theta)$:

$$\hat{\theta} = \arg\max_{\theta} \int \int \int p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) \, d\mathbf{u} \, d\tilde{\mathbf{A}} \, d\mathbf{b}. \tag{26}$$

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. To analyze this further, even having an estimation of the model parameters, we could follow the Bayesian paradigm to estimate the model variables $\mathbf{h} = \{\mathbf{u}_x, \mathbf{u}_y, \mathbf{A}_x, \mathbf{A}_y, \mathbf{b}\}$ by taking their expectation with respect to their posterior:

$$\hat{\mathbf{h}} = \int \mathbf{h} p(\mathbf{h}|\mathbf{d}) \, d\mathbf{h}, \tag{27}$$

where the posterior $p(\mathbf{h}|\mathbf{d})$ is given by Bayes rule:

$$p(\mathbf{h}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{h}) \, p(\mathbf{h})}{p(\mathbf{d})}. \tag{28}$$

However, in our case, as in most models of interest, the previous integral in (27) is intractable.

Thus, we resort to the variational methodology [5], [8] 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* [8]. More specifically, according to the variational Bayesian methodology, we can obtain an approximate posterior for the unknown (hidden) model variables $\mathbf{h}$, by maximizing the bound:

$$L(\mathbf{h}; \theta) = KL(p||q) = \log p(\mathbf{d}; \theta) - \int q(\mathbf{h}) \log \frac{q(\mathbf{h})}{p(\mathbf{h}|\mathbf{d}; \theta)} d\mathbf{h}, \tag{29}$$

with respect to q and θ. As it can be seen, the bound is obtained by subtracting from the marginal log-likelihood $p(\mathbf{d}; \theta)$ the (always positive) Kullback-Leibler (KL) divergence between an arbitrary distribution q and the posterior of $\mathbf{h}$:

$$KL(p||q) = \int q(\mathbf{h}) \log \frac{q(\mathbf{h})}{p(\mathbf{h}|\mathbf{d}; \theta)} d\mathbf{h}. \tag{30}$$

Here, q plays the role of the approximate inferred posterior of $\mathbf{h}$, which is used as a surrogate of the actual posterior for inference purposes. Specifically, since inference using the actual posterior is intractable, we adopt the *mean-field* approximation, where we assume that $\mathbf{u}_x$, $\mathbf{u}_y$, $\mathbf{d}$, $\mathbf{A}_x$ and $\mathbf{A}_y$ are independent in the inferred posterior:

$$q(\mathbf{h}) = \prod_{l=1}^{5} q(\mathbf{h}_l) = q(\mathbf{u}_x) q(\mathbf{u}_y) q(\mathbf{b}) q(\mathbf{A}_x)(\mathbf{A}_y), \tag{31}$$

where $\mathbf{h}_l$ is one of the five variables of the set $\mathbf{h}$. This assumption, although desirable for bypassing the intractability, obviously constrains q from becoming equal to the posterior, whatever the choice of each $q(\mathbf{h}_l)$.

Hence, inference is performed by maximizing the bound L with respect to θ and q iteratively, using the assumption described in (31). The update equations are given abstractly by [8], [4]:

$$\text{VE-step:} \quad q^{(t)}(\mathbf{h}_l) = \underset{q(\mathbf{h}_l)}{\operatorname{argmin}} \, L(q(\mathbf{h}), \theta^{(t-1)}), \; \forall l, \tag{32}$$

$$\text{VM-step:} \quad \theta^{(t)} = \underset{\theta}{\operatorname{argmin}} \, L(q^{(t)}(\mathbf{h}), \theta), \tag{33}$$

where t is the iteration number. The general case update equations that solve are given by the following equations [4]:

$$q^{(t)}(\mathbf{h}_l) = \frac{\exp\left(\log\langle p(\mathbf{d}, \mathbf{h})\rangle_{q^{(t-1)}(\mathbf{h}_{\backslash l})}\right)}{\int_{\mathbf{h}_l} \exp\left(\log\langle p(\mathbf{d}, \mathbf{h})\rangle_{q^{(t-1)}(\mathbf{h}_{\backslash l})}\right) d\mathbf{h}_l}, \tag{34}$$

where $\mathbf{h}_{\backslash l}$ denotes the set $\mathbf{h}$, excluding $\mathbf{h}_l$. Hence, with the mean-field approximation, the iterative scheme of (32) and (33) is a variational EM (VEM) algorithm, which performs approximate inference. It would be an EM algorithm, thus performing exact inference, when we could set $q^{(t)}(\mathbf{h}) = q^{(t)}(\mathbf{h}|\mathbf{d})$ in the VE-step.

In summary, we can obtain an insight of the VEM algorithm by noticing, that in the VE-step, the bound L becomes tight, i.e. it approximates the marginal log-likelihood. Then, in the VM-step, the quantity close to the marginal log-likelihood is maximized with respect to the parameters θ. Thus, since the bound is not exact, this procedure provides a solution close to the maximum likelihood (ML) solution, where the latter could be achieved if the bound was made exact in the first step.

Therefore, in the variational E-step (VE-step) of the algorithm, the posteriors are inferred for all hidden variables using (34). The VE-step is described analytically in the Appendix. Specifically for the motion vectors, the posteriors are Gaussian:

$$q^{(t+1)}(\mathbf{u}_x) = \mathcal{N}\left(\mathbf{m}_x^{(t+1)}, \mathbf{R}_x^{(t+1)}\right), \tag{35}$$

and

$$q^{(t+1)}(\mathbf{u}_y) = \mathcal{N}\left(\mathbf{m}_y^{(t+1)}, \mathbf{R}_y^{(t+1)}\right), \tag{36}$$

where

$$\mathbf{u}_x^{(t+1)} = \mathbf{m}_x^{(t+1)} = \lambda_{noise}^{(t)} \mathbf{R}_x^{(t)} \mathbf{B}^{(t)} \mathbf{G}_x \left(\mathbf{d} - \mathbf{G}_y \mathbf{u}_y^{(t)}\right), \tag{37}$$

and

$$\mathbf{u}_y^{(t+1)} = \mathbf{m}_y^{(t+1)} = \lambda_{noise}^{(t)} \mathbf{R}_y^{(t)} \mathbf{B}^{(t)} \mathbf{G}_y \left(\mathbf{d} - \mathbf{G}_x \mathbf{u}_x^{(t)}\right). \tag{38}$$

Also

$$\mathbf{R}_x^{(t+1)} = \left(\lambda_{noise}^{(t)} \mathbf{G}_x^T \mathbf{B}^{(t)} \mathbf{G}_x + \lambda_x^{(t)} \mathbf{Q}^T \mathbf{A}_x^{(t)} \mathbf{Q}\right)^{-1}, \tag{39}$$

and

$$\mathbf{R}_y^{(t+1)} = \left(\lambda_{noise}^{(t)} \mathbf{G}_y^T \mathbf{B}^{(t)} \mathbf{G}_y + \lambda_y^{(t)} \mathbf{Q}^T \mathbf{A}_y^{(t)} \mathbf{Q}\right)^{-1}. \tag{40}$$

Moreover, the posteriors for $\boldsymbol{\alpha}_x(i)$, $\boldsymbol{\alpha}_y(i)$ and $\mathbf{b}(i)$ are Gamma distributions:

$$q^{(t+1)}(\boldsymbol{\alpha}_k(i)) = \text{Gamma}\left(\frac{\nu_k^{(t)}}{2} + \frac{1}{2}, \frac{\nu_k^{(t)}}{2} + \frac{1}{2}\lambda_k^{(t)}\left(\left[\mathbf{Qu}_k^{(t)}\right]_i^2 + \mathbf{C}_k^{(t)}(i,i)\right)\right), \qquad (41)$$

and

$$q^{(t+1)}(\mathbf{b}(i)) = \text{Gamma}\left(\frac{\mu^{(t)}}{2} + \frac{1}{2}, \frac{\mu^{(t)}}{2} + \frac{1}{2}\lambda_{noise}^{(t)}\left(\left[\mathbf{Gu}^{(t)} - \mathbf{d}\right]_i^2 + \mathbf{F}^{(t)}(i,i)\right)\right), \qquad (42)$$

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

$$\langle \boldsymbol{\alpha}_k(i) \rangle = \frac{\nu_k^{(t)} + 1}{\nu_k^{(t)} + \lambda_k^{(t)}\left(\left[\mathbf{Qu}_k^{(t)}\right]_i^2 + \mathbf{C}_k^{(t)}(i,i)\right)}, \qquad (43)$$

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

$$\mathbf{C}_k^{(t)} = \mathbf{Q}\mathbf{R}_k^{(t)}\mathbf{Q}^T. \qquad (44)$$

Notice that $\boldsymbol{\alpha}_k(i)$ is the equivalent parameter present in many variational methods [18], [12] 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:

$$\langle \mathbf{b}(i) \rangle = \frac{\mu^{(t)} + 1}{\mu^{(t)} + \lambda_{noise}^{(t)}\left(\left[\mathbf{Gu}^{(t)} - \mathbf{d}\right]_i^2 + \mathbf{F}^{(t)}(i,i)\right)}, \qquad (45)$$

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

$$\mathbf{F}^{(t)} = \mathbf{G}_x\mathbf{R}_x^{(t)}\mathbf{G}_x^T + \mathbf{G}_y\mathbf{R}_y^{(t)}\mathbf{G}_y^T. \qquad (46)$$

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 (43) and (45) 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 [29] for their calculation. For matrices $\mathbf{C}_x$, $\mathbf{C}_y$ and $\mathbf{F}$ only the diagonal elements are needed in (43) and (45) and they are obtained as a byproduct of the Lanczos method.

Let us notice that as we can see from (37) and (38), 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:

$$\lambda_{noise}^{(t+1)} = \frac{N}{\sum_{i=1}^{N} \langle \mathbf{b}(i) \rangle \left(\left[\mathbf{Gu}^{(t+1)} - \mathbf{d} \right]_i^2 + \mathbf{F}^{(t+1)}(i,i) \right)}, \qquad (47)$$

and equivalently for λ_x and λ_y:

$$\lambda_k^{(t+1)} = \frac{N}{\sum_{i=1}^{N} \langle \boldsymbol{\alpha}_k(i) \rangle \left(\left[\mathbf{Qu}_k^{(t+1)} \right]_i^2 + \mathbf{C}_k^{(t+1)}(i,i) \right)}, \qquad (48)$$

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:

$$\frac{1}{N} \left(\sum_{i=1}^{N} \log\langle \boldsymbol{\alpha}_k(i) \rangle - \sum_{i=1}^{N} \langle \boldsymbol{\alpha}_k(i) \rangle \right) + F\left(\frac{\nu_k^{(t)}}{2} + \frac{1}{2} \right) - \log\left(\frac{\nu_k^{(t)}}{2} + \frac{1}{2} \right) - F\left(\frac{\nu_k}{2} \right) + \log\left(\frac{\nu_k}{2} \right) + 1 = 0, \qquad (49)$$

for ν_k, $k \in \{x, y\}$, where $F(x)$ 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

$$\frac{1}{N} \left(\sum_{i=1}^{N} \log\langle \mathbf{b}(i) \rangle - \sum_{i=1}^{N} \langle \mathbf{b}(i) \rangle \right) + F\left(\frac{\mu^{(t)}}{2} + \frac{1}{2} \right) - \log\left(\frac{\mu^{(t)}}{2} + \frac{1}{2} \right) - F\left(\frac{\mu}{2} \right) + \log\left(\frac{\mu}{2} \right) + 1 = 0. \qquad (50)$$

In our implementation, equations (49) and (50) are solved by the bisection method, as also proposed in [23]. 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 [18].

V. EXPERIMENTAL RESULTS

The method proposed herein is a principled Bayesian generalization of the Horn-Schunck (HS) method [18]. Therefore, our purpose is to examine its appropriateness to replace it in the initialization of more advanced optical flow schemes. We have also included the well-known and established rival algorithm of Lucas-Kanade (LK) [24]. 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 [7], which combines the above two algorithms for feature tracking, based on a framework proposed in [13]. We call this method *Joint Lucas-Kanade* (JLK). To visualize the motion vectors we adopt the color coding fig. 2.

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

---
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 (43).
        * Compute the expectation $\langle \mathbf{b}(i) \rangle$ using (45).
    - **VM-step:**
        * Compute λ_{noise} using (47).
        * Compute λ_x, λ_y using (48).
        * Solve for ν_x, ν_y equation (49), using the bisection method.
        * Solve for μ equation (50), using the bisection method.
        * Update matrices $\mathbf{R}_x$ and $\mathbf{R}_y$ in (39) and (40) and the diagonal elements of $\mathbf{C}_x$, $\mathbf{C}_y$ and $\mathbf{F}$ using the Lanczos method.
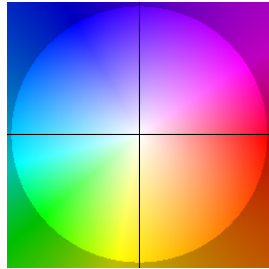   * Compute $\mathbf{u}_x$, $\mathbf{u}_y$ from (37) and (38).

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

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° 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] and other reliable resources, in cases where the Middlebury database does not provide the data we need. More specifically, since the ground truth data for the (Yosemite) optical flow case is not available by the Middlebury web-site, we were obliged to obtain these data from other reliable sources. As a consequence, there is not a clear correspondence between the experiments we present and the experimental evaluations available at the Middlebury web-site.

More specifically, since the ground truth data for the (Yosemite) optical flow case is not available by the Middlebury web-site, we were obliged to obtain these data from other reliable sources.

In order to evaluate the performance of the method, two performance indices were computed. The first index of performance that we use in the comparison is the average angular error (AAE) [3]. This is the most common measure of performance for optical flow [2]. Let $\mathbf{v}_0 = (u_0 , v_0)$ be the correct velocity and $\mathbf{v}_1 = (u_1 , v_1)$ be the estimated velocity. The angular error (AE) between these two vectors is

$$\psi_{AE} = arccos \left(\vec{\mathrm{v_0}} \cdot \vec{\mathrm{v_1}} \right) \tag{51}$$

where $\vec{\mathrm{v_0}}$, $\vec{\mathrm{v_1}}$ are the 3D normalized representations of $\mathbf{v}_0$, $\mathbf{v}_1$, respectively and defined as

$$\vec{\mathrm{v_0}} = \frac{1}{\sqrt{u_0{}^2 + v_0{}^2 + 1}} (u_0, v_0, 1) \tag{52}$$

$$\vec{\mathrm{v_1}} = \frac{1}{\sqrt{u_1{}^2 + v_1{}^2 + 1}} (u_1, v_1, 1) \tag{53}$$

The AAE is then obtained by calculating the average of all angular errors between correct and estimated velocities in the optical flow. However, it can be seen from Eq. (51) that errors in regions of large flows are penalized less in AE than errors in regions of small flows [2]. Thus, one needs to be cautious when using the AAE metric as estimates with the same error magnitude may result in significantly different angular error values.

Another error metric is the normalized magnitude of the vector difference between the correct and estimated flow vectors [25]. The magnitude of the correct velocity is used as a normalization factor. The magnitude of difference error is defined as

$$E_M = \begin{cases} \frac{\|\mathbf{v}_0 - \mathbf{v}_1\|}{\|\mathbf{v}_0\|} & , \text{if } \|\mathbf{v}_0\| \geq T \\ \left| \frac{\|\mathbf{v}_1\| - T}{T} \right| & , \text{if } \|\mathbf{v}_0\| < T \text{ and } \|\mathbf{v}_1\| \geq T \\ 0 & , \text{if } \|\mathbf{v}_0\| < T \text{ and } \|\mathbf{v}_1\| < T \end{cases} \tag{54}$$

where T is a threshold, whose purpose is to normalize the errors with respect to the ground truth and to ignore normalized vector norms with values less than T. An algorithm is not expected to reliably produce accurate flow vectors in areas where the actual flow magnitude is less than the value of parameter T [25]. We used $T = 0.35$ in all of our experiments. The average magnitude of difference error (AME) is then calculated as the average of the normalized magnitude of difference errors.

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 [7] but not for dense flow estimation.

Representative results are presented in fig. 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 [24]	5.91°	0.15	8.58°	0.17	11.65°	0.26	27.52°	0.56
Horn-Schunck [18]	2.47°	0.05	5.57°	0.14	5.43°	0.13	8.50°	0.49
JLK [7]	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 and AME for all of the compared algorithms, shown in fig. 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 [18] 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 is a clear gain in accuracy.

Moreover, in order to assess the robustness and consistency of the proposed algorithm with respect to the parameter estimation procedure, we conducted three experimental setups. We kept fixed a subset of the parameters and we let the algorithm estimate the rest (with the standard update scheme of the proposed algorithm). More specifically, we used the following three parameter sets:
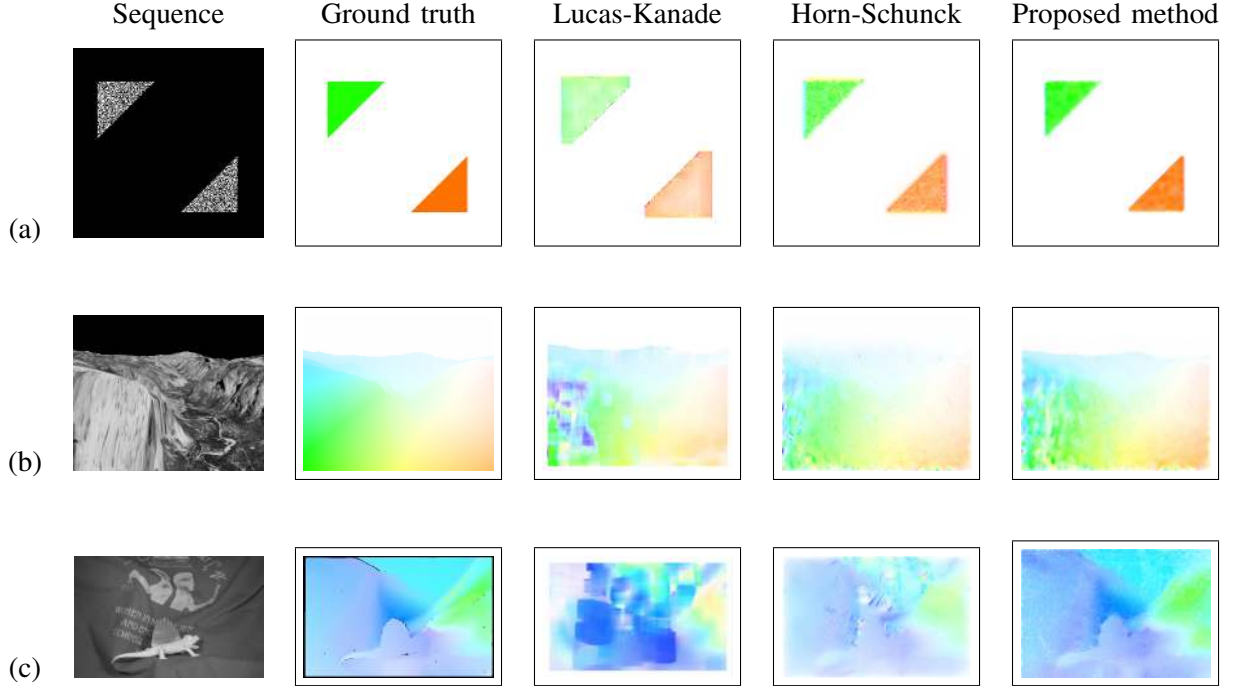
- μ

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

- ν_x, ν_y,
- ν_x, ν_y and m.

The parameters are kept fixed to the value that provides the best performance, in terms of the AAE and and AME metrics. These values were found by a trial-end-error procedure. The results of these experiments are presented in Table II. From this table, we can conclude that, although we fix the parameters to an optimal value, the results are not significantly better than the proposed algorithm, where all parameters are found automatically. This demonstrates the robustness of the proposed algorithm with regard to parameter estimation.

In order to justify our assumption that the spatial differences ϵ of the optical flow vectors $(\mathbf{u}_x, \mathbf{u}_y)$ may be modeled as random variables following a Student's-*t* distribution, we conducted experiments by fitting a Student's-*t* and a Gaussian distribution to the local differences of the ground truth flow vectors, obtained from the Middlebury database. More specifically, we used the EM algorithm to estimate the parameters of the distributions by maximizing the log-likelihood of the motion field data with respect to these parameters. The values of the log-likelihood at convergence, are given in Table III, where it may be observed that the Student's *t*-distribution provides more accurate modeling of the motion field differences than the Gaussian in most cases. Note that all log-likelihood values are multiples of 10^5. Moreover, in fig. 5, the histograms of the horizontal and vertical differences of
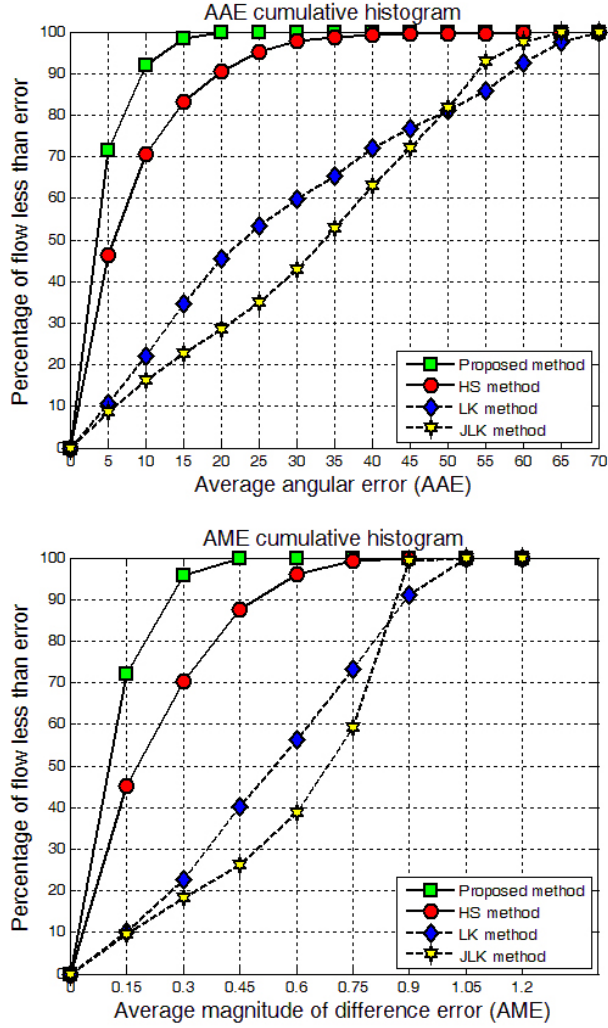
Fig. 4. Performances of the compared algorithms on the *Dimetrodon* sequence [2]. Cumulative histograms showing the percentage of the optical flow errors which are lower than a given value (represented along the horizontal axis) for the AAE (top) and the AME (bottom).

the optical flow for the *Yosemite* sequence are depicted, based on ground truth data. On the same figure, the fitting of a Student's-*t* and a Gaussian distribution are also shown. As it can be deduced, the Student's *t*-distribution provides a better fitting (and thus modeling) than the Gaussian. Note that the zero-mean assumption of the local differences of $\mathbf{u}_x$ and $\mathbf{u}_y$ is also verified by the histograms in fig. 5. Also, in Table IV, analogous results for testing the modeling accuracy of the observation vector d are given. The experimental results support our choice of modeling the noise vector $\mathbf{d}$ with a Student's *t*- distribution, instead of a Gaussian, also in this case.

The algorithm takes on average less than a minute to converge on 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

TABLE II

OPTICAL FLOW ERRORS OF THE PROPOSED ALGORITHM, WHERE SOME PARAMETERS ARE KEPT FIXED IN AN OPTIMAL

WAY (IN TERMS OF AAE AND AME).

Fixed parameters	Triangles-Equal		Triangles-Unequal		Yosemite	
	AAE	AME	AAE	AME	AAE	AME
$\mu = 0.5$	$1.008°$	0.024	**3.90°**	**0.087**	$4.44°$	0.11
$\mu = 1.5$	$1.006°$	0.024	$3.92°$	0.093	**4.42°**	**0.11**
$\nu_x = 5.5, \nu_y = 7.5$	$1.06°$	0.024	$3.924°$	0.095	**4.44°**	**0.11**
$\nu_x = 10.5, \nu_y = 8.5$	**1.04°**	**0.024**	$3.93°$	0.096	$4.46°$	0.11
$\nu_x = 6.5, \nu_y = 9.5$	$1.05°$	0.024	**3.917°**	**0.095**	$4.46°$	0.11
$\nu_x = 9.5, \nu_y = 7.5, \mu = 0.5$	$1.007°$	0.024	**3.89°**	**0.086**	$4.446°$	0.11
$\nu_x = 9.5, \nu_y = 8.5, \mu = 1.5$	$1.00°$	0.024	$1.00°$	0.024	**4.419°**	**0.11**
$\nu_x = 10.5, \nu_y = 5.5, \mu = 1.5$	**1.00°**	**0.024**	$3.91°$	0.093	$4.423°$	0.11
None (standard algorithm)	$1.06°$	0.024	$3.93°$	0.0957	$4.45°$	0.11

TABLE III

LOG-LIKELIHOOD VALUES ($\times 10^5$) OBTAINED BY FITTING A GAUSSIAN AND A STUDENT'S t-DISTRIBUTION TO THE

LOCAL DIFFERENCES OF THE GROUND TRUTH MOTION-FIELD DATA.

	Dimetrodon		Grove2		Grove3	
	$\mathbf{u}_x$	$\mathbf{u}_y$	$\mathbf{u}_x$	$\mathbf{u}_y$	$\mathbf{u}_x$	$\mathbf{u}_y$
Gaussian	3.89	6.01	8.24	2.79	2.84	0.64
Student's t	4.61	40.77	14.82	12.54	13.66	14.56

	Hydrangea		Rubberwhale		Urban2	
	$\mathbf{u}_x$	$\mathbf{u}_y$	$\mathbf{u}_x$	$\mathbf{u}_y$	$\mathbf{u}_x$	$\mathbf{u}_y$
Gaussian	0.40	2.18	3.65	5.56	0.96	4.64
Student's t	5.95	5.97	5.36	5.54	11.20	14.70

	Urban3		Yosemite without clouds		Yosemite with clouds	
	$\mathbf{u}_x$	$\mathbf{u}_y$	$\mathbf{u}_x$	$\mathbf{u}_y$	$\mathbf{u}_x$	$\mathbf{u}_y$
Gaussian	5.33	4.64	1.63	1.81	1.15	1.83
Student's t	24.50	15.10	24.70	30.31	24.17	28.14

in Eq. (39)-(40).

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-
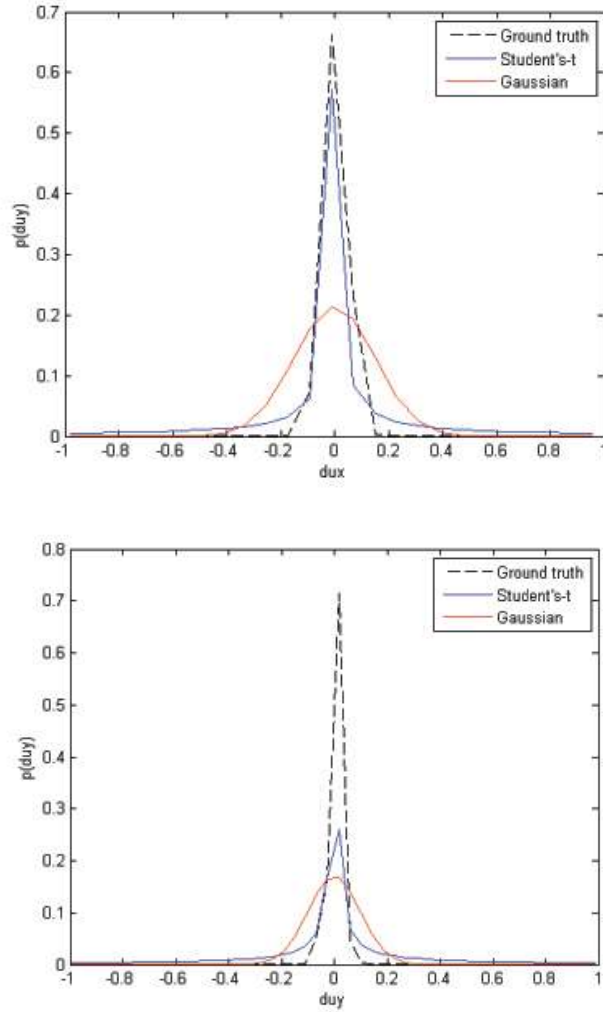
Fig. 5. Fitting the ground truth motion field histogram of the *Yosemite* sequence with a Student's t-distribution and a Gaussian distribution, using the EM algorithm. du_x and du_y are the local differences of the vertical and horizontal motion field respectively. Note that the motion-field values have been divided by 256.

t distribution of the motion vectors achieves selective application of smoothness leaving motion edges unaffected. This is due to the heavy-tailed property of the Student's *t*-distribution, which provides robust motion field estimates. Furthermore, any residuals of the linearization of the brightness constancy constraint are also modeled leading to better accuracy. Moreover, the robustness of the proposed algorithm and the superiority of Student's *t*-distribution with respect to the Gaussian for the optical flow modeling was demonstrated.

The proposed method is a variational formulation of the Horn-Schunck method, which is widely used as an initialization module in other motion estimation algorithms. The Horn-Schunck method considers only small motions and therefore it cannot account for large motion fields. From this point of view, the approach described herein may be seen as an important contribution to the process

TABLE IV

LOG-LIKELIHOOD VALUES ($\times 10^5$) OBTAINED BY FITTING A GAUSSIAN AND A STUDENT'S t-DISTRIBUTION TO THE OBSERVATION VECTOR $\mathbf{d}$.

	Dimetrodon	*Grove2*	*Grove3*
Gaussian	5.48	2.5	3.36
Student's t	6.61	2.74	4.00

	Hydrangea	*Rubberwhale*	*Urban2*
Gaussian	3.97	5.70	5.41
Student's t	4.81	6.19	6.43

	Urban3	*Yosemite without clouds*	*Yosemite with clouds*
Gaussian	5.45	1.46	1.40
Student's t	6.26	1.73	1.66

of initializing optical flow estimators. Moreover, the proposed variational-Bayesian method may be considered jointly with other variational methods incorporating more sophisticated constraints on the motion field in order to formulate them in a probabilistic framework. More specifically, the methods in [12] and [13], also based on the Horn-Schunck assumption, could follow the same stochastic principles.

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. Variational algorithms for approximate Bayesian inference. Technical report, The Gatsby Computational Neuroscience Unit, University College, London, 2003.

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

[6] 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.

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

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

[9] 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.

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

[11] M. Black and P.Anandan. A framework for the robust estimation of optical flow. In *Fourth International Conference on Computer Vision*, pages 231–236, May 1993.

[12] 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.

[13] 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.

[14] 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.

[15] T. Gkamas, G. Chantas, and C. Nikou. A probabilistic formulation of the optical flow problem. In 21st *International Conference on Pattern Recognition (ICPR)*, Tsukuba, Japan, 11-15 November 2012.

[16] 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.

[17] P. Heas, C. Herzet, and E. Memin. Robust optic-flow estimation with bayesian inference of model and hyper-parameters. *Scale Space and Variational methods in Computer Vision, LNCS*, 6667:773–785, 2012.

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

[19] K. Krajsek and A. Mester. Bayesian inference and maximum entropy methods in science and engineering, aip conference proceedings. *Scale Space and Variational methods in Computer Vision, LNCS*, 872:311–318, 2006.

[20] K. Krajsek and A. Mester. Bayesian model selection for optical flow estimation. In *Proceedings of the 29th DAGM Conference on Pattern recognition*, Heidelberg, 2007.

[21] K. Krajsek and R. Mester. A maximum likelihood estimator for choosing the regularization parameters in global optical flow methods. In *IEEE International Conference on Image Processing*, pages 1081–1084, 2006.

[22] 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.

[23] 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.

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

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

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

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

[32] G. Sfikas, C. Heinrich, J. Zallat, C. Nikou, and N. Galatsanos. Recovery of polarimetric Stokes images by spatial

mixture models. *Journal of the Optical Society of America A*, 28(3):465–474, 2011.

[33] G. Sfikas, C. Nikou, and N. Galatsanos. Edge preserving spatially varying mixtures for image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, Alaska, USA, 2008.

[34] G. Sfikas, C. Nikou, N. Galatsanos, and C. Heinrich. Spatially varying mixtures incorporating line processes forimage segmentation. *Journal of Mathematical Imaging and Vision*, 36:91–110, 2010.

[35] G. Sfikas, C. Nikou, N. Galatsanos, and C. Heinrich. Majorization-minimization mixture model determination in image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2169–2176, Colorado Springs, Colorado, USA, 2011.

[36] 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.

[37] D. Tzikas, A. Likas, and N. Galatsanos. The variational approximation for Bayesian inference. *IEEE Signal Processing Magazine*, 25(6):131 –146, 2008.

[38] 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.

[39] 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.

[40] 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.

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

[42] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:1744–1757, 2012.

[43] Z. Zhou, R. M. Leahy, and J. Qi. Approximate maximum likelihood hyperparameter estimation for gibbs priors. *IEEE Transactions on Image Processing*, 6(6):844–861, June 1997.

[44] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, 2011.

[45] 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.

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:

$$p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) = p\left(\mathbf{d}|\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\mathbf{u}|\tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\tilde{\mathbf{A}}; \theta\right) p\left(\mathbf{b}; \theta\right), \tag{55}$$

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 $p\left(\mathbf{d}; \theta\right)$:

$$\hat{\theta} = \arg\max_{\theta} \int \int \int p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) d\mathbf{u} d\tilde{\mathbf{A}} d\mathbf{b}. \tag{56}$$

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 [5], [8] where we have to maximize a lower bound of $p\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right)$:

$$L\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) = \int \int \int q\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right) \log \frac{p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right)}{q\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right)} d\mathbf{u} d\tilde{\mathbf{A}} d\mathbf{b}. \tag{57}$$

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 (57) becomes equality. However, in the variational methodology, we employ the *mean field* approximation [8]:

$$q\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right) = q\left(\mathbf{u}\right) q\left(\tilde{\mathbf{A}}\right) q\left(\mathbf{b}\right), \tag{58}$$

and (57) becomes:

$$L\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) = \int_{\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}} q\left(\mathbf{u}\right) q\left(\tilde{\mathbf{A}}\right) q\left(\mathbf{b}\right) \log \frac{p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right)}{q\left(\mathbf{u}\right) q\left(\tilde{\mathbf{A}}\right) q\left(\mathbf{b}\right)} \tag{59}$$

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:

$$q\left(\mathbf{u}\right) = \mathcal{N}\left(\left[\begin{array}{c} \mathbf{m}_x \\ \mathbf{m}_y \end{array}\right], \left[\begin{array}{cc} \mathbf{R}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_y \end{array}\right]\right), \tag{60}$$

yielding

$$q\left(\mathbf{u}_x\right) = \mathcal{N}\left(\mathbf{m}_x, \mathbf{R}_x\right), \tag{61}$$

and

$$q\left(\mathbf{u}_y\right) = \mathcal{N}\left(\mathbf{m}_y, \mathbf{R}_y\right). \tag{62}$$

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 (58), the variational bound in our problem becomes:

$$L\left(q(\mathbf{u}_x), q(\mathbf{u}_y), q(\mathbf{a}_x), q(\mathbf{a}_y), q(\mathbf{b}), \theta_1, \theta_2\right) =$$

$$\int \int \int \left(\prod_{k \in \{x,y\}} q(\mathbf{u}_k; \theta_1) q(\mathbf{a}_k)\right) q(\mathbf{b}) \log p(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta_2) d\mathbf{u}\ d\tilde{\mathbf{A}} d\mathbf{b}$$

$$- \int \int \int \left(\prod_{k \in \{x,y\}} q(\mathbf{u}_k; \theta_1) q(\mathbf{a}_k)\right) q(\mathbf{b}) \log \left(\left(\prod_{k \in \{x,y\}} p(\mathbf{u}_k; \theta_1) q(\mathbf{a}_k)\right) q(\mathbf{b})\right) d\mathbf{u}\ d\tilde{\mathbf{A}} d\mathbf{b} \tag{63}$$

where we have separated the parameters into two sets:

$$\theta_1 = \{\mathbf{R}_x, \mathbf{R}_x, \mathbf{m}_x, \mathbf{m}_y\}, \tag{64}$$

and

$$\theta_2 = \{\mathbf{a}_x, \mathbf{a}_y, \mathbf{b}, \lambda_x, \lambda_y, \nu_x, \nu_y\}. \tag{65}$$

Thus, in the VE-step of the variational EM 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 (63) 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:

$$\mathbf{m}_x^{(t+1)} = \lambda_{noise}^{(t)} \mathbf{R}_x^{(t)} \hat{\mathbf{B}}^{(t)} \mathbf{G}_x \left(\mathbf{d} - \mathbf{G}_y \mathbf{u}_y^{(t)} \right), \tag{66}$$

and

$$\mathbf{m}_y^{(t+1)} = \lambda_{noise}^{(t)} \mathbf{R}_y^{(t)} \hat{\mathbf{B}}^{(t)} \mathbf{G}_y \left(\mathbf{d} - \mathbf{G}_x \mathbf{u}_x^{(t)} \right), \tag{67}$$

where

$$\mathbf{R}_x^{(t+1)} = \left(\lambda_{noise}^{(t)} \mathbf{G}_x^T \hat{\mathbf{B}}^{(t)} \mathbf{G}_x + \lambda_x^{(t)} \mathbf{Q}^T \hat{\mathbf{A}}_x^{(t)} \mathbf{Q} \right)^{-1}, \tag{68}$$

and

$$\mathbf{R}_y^{(t+1)} = \left(\lambda_{noise}^{(t)} \mathbf{G}_y^T \hat{\mathbf{B}}^{(t)} \mathbf{G}_y + \lambda_y^{(t)} \mathbf{Q}^T \hat{\mathbf{A}}_y^{(t)} \mathbf{Q} \right)^{-1}, \tag{69}$$

Notice that the final estimates for $\mathbf{u}_x$, $\mathbf{u}_y$ are $\mathbf{m}_x$ and $\mathbf{m}_y$, in (37) and (38), respectively.

After some manipulation, we obtain the update equations for the model parameters which maximize (63) with respect to $q(\mathbf{a}_k)$, $q(\mathbf{b})$. 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 $q(\mathbf{a}_k)$, $q(\mathbf{b})$ which approximate $p(\mathbf{a}_k|\mathbf{u}_k, \lambda_k, \mathbf{C}_k; \nu_k)$, $p(\mathbf{b}|\mathbf{u}, \lambda_{noise}, \mathbf{F}; \mu)$ will follow Gamma distributions, $\forall i = 1, ..., N, \forall k \in \{x, y\}$:

$$q^{(t+1)}(\boldsymbol{\alpha}_k(i)) = \text{Gamma}\left(\frac{\nu_k^{(t)}}{2} + \frac{1}{2}, \frac{\nu_k^{(t)}}{2} + \frac{1}{2}\lambda_k^{(t)} \left(\left[\mathbf{Q}\mathbf{u}_k^{(t)} \right]_i^2 + \mathbf{C}_k^{(t)}(i,i) \right) \right), \tag{70}$$

and

$$q^{(t+1)}(\mathbf{b}(i)) = \text{Gamma}\left(\frac{\mu^{(t)}}{2} + \frac{1}{2}, \frac{\mu^{(t)}}{2} + \frac{1}{2}\lambda_{noise}^{(t)} \left(\left[\mathbf{G}\mathbf{u}^{(t)} - \mathbf{d} \right]_i^2 + \mathbf{F}^{(t)}(i,i) \right) \right), \tag{71}$$

where the $N \times N$ matrix

$$\mathbf{C}_k^{(t)} = \mathbf{Q}\mathbf{R}_k^{(t)}\mathbf{Q}^T, \tag{72}$$

the $N \times N$ matrix

$$\mathbf{F}^{(t)} = \mathbf{G}_x \mathbf{R}_x^{(t)} \mathbf{G}_x^T + \mathbf{G}_y \mathbf{R}_y^{(t)} \mathbf{G}_y^T, \tag{73}$$

and $[\cdot]_i$ denotes the i-th element of the vector inside the brackets.

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 [29] for their calculation. For matrices $\mathbf{C}_x$, $\mathbf{C}_y$ and $\mathbf{F}$ only the diagonal elements are needed in (70) and (71) and they are obtained as a byproduct of the Lanczos method.

Note that as we can see from (66) and (67), 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:

$$\langle \boldsymbol{\alpha}_k(i)\rangle_{q^{(t+1)}(\boldsymbol{\alpha}_k(i))} = \frac{\nu_k^{(t)}+1}{\nu_k^{(t)}+\lambda_k^{(t)}\left(\left[\mathbf{Q}\mathbf{u}_k^{(t)}\right]_i^2+\mathbf{C}_k^{(t)}(i,i)\right)}, \tag{74}$$

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

$$\langle \mathbf{b}(i)\rangle_{q^{(t+1)}(\mathbf{b}(i))} = \frac{\mu^{(t)}+1}{\mu^{(t)}+\lambda_{noise}^{(t)}\left(\left[\mathbf{G}\mathbf{u}^{(t)}-\mathbf{d}\right]_i^2+\mathbf{F}^{(t)}(i,i)\right)}, \tag{75}$$

where $\langle.\rangle_{q(.)}$ denotes the expectation with respect to an arbitrary distribution $q(\cdot)$. These estimates are used in (66), (67), (68) and (69), 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:

$$\theta_2^{(t+1)} = \arg\max_{\theta_2} L\left(q^{(t+1)}\left(\mathbf{u}_k\right), q^{(t+1)}\left(\hat{\mathbf{A}}_k\right), q^{(t+1)}\left(\hat{\mathbf{B}}\right), \theta_1^{(t+1)}, \theta_2\right), \tag{76}$$

where

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

is calculated using the results from (66) - (69).

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 (63) with respect to it and setting it to zero:

$$\lambda_{noise}^{(t+1)} = \frac{N}{\sum_{i=1}^{N} \mathbf{b}^{(t+1)}(i)\left(\left[\mathbf{G}\mathbf{u}^{(t+1)}-\mathbf{d}\right]_i^2+\mathbf{F}^{(t+1)}(i,i)\right)}. \tag{77}$$

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

$$\lambda_k^{(t+1)} = \frac{N}{\sum_{i=1}^{N} \boldsymbol{\alpha}_k^{(t+1)}(i)\left(\left[\mathbf{Q}\mathbf{u}_k^{(t+1)}\right]_i^2+\mathbf{C}_k^{(t+1)}(i,i)\right)}, \tag{78}$$

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:

$$\frac{1}{N}\left(\sum_{i=1}^{N}\log\langle\boldsymbol{\alpha}_k(i)\rangle_{q^{(t+1)}(\mathbf{A}_k)} - \sum_{i=1}^{N}\langle\boldsymbol{\alpha}_k(i)\rangle_{q^{(t+1)}(\mathbf{A}_k)}\right) + F\left(\frac{\nu_k^{(t)}}{2}+\frac{1}{2}\right)$$

$$- \log\left(\frac{\nu_k^{(t)}}{2}+\frac{1}{2}\right) - F\left(\frac{\nu_k}{2}\right) + \log\left(\frac{\nu_k}{2}\right) + 1 = 0, \tag{79}$$

for ν_k, $k \in \{x, y\}$, where $F(x)$ 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:

$$\frac{1}{N}\left(\sum_{i=1}^{N}\log\langle\mathbf{b}(i)\rangle_{q^{(t+1)}(\mathbf{b}(i))} - \sum_{i=1}^{N}\langle\mathbf{b}(i)\rangle_{q^{(t+1)}(\mathbf{b}(i))}\right) +$$

$$F\left(\frac{\mu^{(t)}}{2}+\frac{1}{2}\right) - \log\left(\frac{\mu^{(t)}}{2}+\frac{1}{2}\right) - F\left(\frac{\mu}{2}\right) +$$

$$\log\left(\frac{\mu}{2}\right) + 1 = 0. \tag{80}$$

In our implementation equations (79) and (80) are solved by the bisection method, as also proposed in [23].

Giannis Chantas received the Diploma, MSc, and PhD degrees from the Department of Computer Science, University of Ioannina, in 2002, 2004 and 2009 respectively. From 2009 to 2011 he has been with the Department of Computer Science, TEI of Larisa. Currently, he holds a postdoctoral position in the Artificial Intelligence and Information Analysis laboratory of the Department of Computer Science of Aristotle University of Thessaloniki. His research interest lies in the area of Bayesian inference for computer vision and image and video processing.

Theodosios Gkamas was born in Ioannina, Greece, in 1985. He obtained the B.S. degree (2008) and the M.Sc. degree (2010) in Computer Science from the Department of Computer Science of the University of Ioannina, Ioannina, Greece. He is currently a Ph.D. candidate of the Graduate School of Mathematics, Information and Engineering Sciences of the University of Strasbourg, Strasbourg, France, where he is a member of the Laboratoire des Sciences de l'Ingenieur, de l'Informatique et de l'Imagerie (iCube), CNRS UMR 7357, research groups MIV and IMIS. His research interests are 3D medical image analysis, machine learning and computer vision.

Christophoros Nikou was born in Thessaloniki, Greece, in 1971. He received the Diploma in electrical engineering from the Aristotle University of Thessaloniki, Greece, in 1994 and the DEA and Ph.D. degrees in image processing and computer vision from Louis Pasteur University, Strasbourg, France, in 1995 and 1999, respectively. He was a Senior Researcher with the Department of Informatics, Aristotle University of Thessaloniki in 2001. From 2002 to 2004, he was a Research Engineer and Project Manager with Compucon S.A., Thessaloniki, Greece. He was a Lecturer (2004-2009) and an Assistant Professor (2009-2013) with the Department of Computer Science and Engineering, University of Ioannina, Ioannina, Greece, where he has been an Assiociate Professor, since 2013. His research interests mainly include image processing and analysis, computer vision and pattern recognition and their application to medical imaging. C. Nikou is a Senior Member of IEEE and a member of EURASIP.