BAYESIAN METHODS FOR OPTICAL FLOW ESTIMATION USING A VARIATIONAL APPROXIMATION, WITH APPLICATIONS TO ULTRASOUND

Jan Dorazil[⋆,†]*, Bernard H. Fleury*[⋆]*, and Franz Hlawatsch*[⋆]

[⋆]Institute of Telecommunications, TU Wien, Vienna, Austria
[†]Department of Telecommunications, Brno University of Technology, Brno, Czech Republic

ABSTRACT

We develop a unified Bayesian framework for optical flow (OF) estimation that uses a variational lower bound to obtain a variational approximation of the posterior probability distribution. Our framework enables the incorporation of domain-specific knowledge as well as a quantification of the uncertainty of OF estimation, and it encompasses existing maximum a posteriori (MAP) and variational Bayes (VB) methods as special cases. We leverage this flexibility for the ultrasound modality by using ultrasound-specific likelihood functions within both MAP and VB methods. Numerical results for the problem of cardiac motion estimation demonstrate that VB methods outperform MAP methods, in addition to providing a more faithful uncertainty measure.

***Index Terms*—** Optical flow, Bayesian estimation, variational approximation, ultrasound, cardiac motion estimation.

1. INTRODUCTION

Optical flow (OF) estimation is an important image processing task with a wide range of applications including video compression, autonomous navigation, and medical diagnosis and intervention [1]. While many recent advances in OF estimation benefit from the use of deep learning methods, model-based methods continue to be of relevance. This is especially true in the medical domain where large training datasets are not widely available and the interpretability of the OF estimation method has a high priority.

Traditionally, OF estimation is formulated as the minimization of a suitably designed "energy" function [2, 3]. From a Bayesian probabilistic viewpoint, this can be interpreted as maximum a posteriori (MAP) estimation involving a posterior probability density function (pdf). This viewpoint led to the development of advanced Bayesian OF estimation methods, typically based on the principle of variational Bayes (VB) [4, 5]. These methods allow for the extraction of additional information from the posterior pdf, including measures of the reliability (uncertainty) of the OF estimate.

Here, we develop a unified Bayesian framework, inspired by [6], that uses a variational lower bound rooted in convex analysis. Our approach differs from the variational lower bound proposed for image restoration in [7] in that we lower-bound a different objective (which is not directly related to the Kullback-Leibler divergence). Our framework encompasses both MAP and VB OF estimation methods and makes it possible to incorporate domain-specific knowledge—in particular, likelihood functions specifically designed for the imaging modality under consideration, rather than standard choices such as the t-distribution [4] or a Gaussian mixture distribution [5]. We leverage this flexibility for the ultrasound modality by using the ultrasound-specific likelihood functions proposed in [8] and [9] within both MAP and VB methods. In addition, our unified framework reveals that also the MAP methods [2], [3] provide a natural way of calculating an uncertainty measure. Numerical results for the problem of cardiac motion estimation from ultrasound sequences demonstrate that VB methods outperform MAP methods, in addition to providing a more faithful uncertainty measure.

The rest of this paper is organized as follows. In Section 2, we describe an energy-based model for OF estimation and a corresponding posterior pdf of the OF field. In Section 3, we present a variational lower bound and use it to obtain tractable approximations to the posterior pdf. In Section 4, we develop two Bayesian OF estimation methods that are based on the variational lower bound. In Section 5, we consider two domain-specific likelihood functions for the ultrasound modality. Finally, in Section 6, we present numerical results for synthetic cardiac ultrasound sequences.

2. BAYESIAN OF ESTIMATION FRAMEWORK

Our setup for OF estimation is based on the well-established brightness constancy principle combined with a smoothness-based regularization; see [2] for a motivation and a more detailed description. Let $F(\boldsymbol{x})$ and $G(\boldsymbol{x})$ with $\boldsymbol{x} = (x\ y)^{\mathrm{T}} \in \mathbb{R}^2$ represent two temporally consecutive images. The OF vector $\boldsymbol{w}(\boldsymbol{x}) = \big(u(\boldsymbol{x})\ v(\boldsymbol{x})\big)^{\mathrm{T}} \in \mathbb{R}^2$ describes the local displacement of the brightness of image $F(\boldsymbol{x})$ at spatial position $\boldsymbol{x}$ when changing to image $G(\boldsymbol{x})$. We consider the samples $\boldsymbol{w}_{\boldsymbol{r}} \triangleq \boldsymbol{w}(\boldsymbol{x}_{\boldsymbol{r}}) = \big(u(\boldsymbol{x}_{\boldsymbol{r}})\ v(\boldsymbol{x}_{\boldsymbol{r}})\big)^{\mathrm{T}} = (u_{\boldsymbol{r}}\ v_{\boldsymbol{r}})^{\mathrm{T}}$ of $\boldsymbol{w}(\boldsymbol{x})$ taken at the vertices $\boldsymbol{x}_{\boldsymbol{r}}$ of a uniform 2D sampling grid, where $\boldsymbol{r} \in \mathcal{R} \triangleq \{1,\dots,N\}\times\{1,\dots,M\}$. It will be convenient to stack all the OF vectors $\boldsymbol{w}_{\boldsymbol{r}}$, $\boldsymbol{r}\in\mathcal{R}$ into an overall OF vector $\boldsymbol{w}$ of length $2NM$. The task is to estimate $\boldsymbol{w}$ from the images $F(\boldsymbol{x})$ and $G(\boldsymbol{x})$.

OF estimation is typically formulated as the minimization of a suitably designed "energy" function [2]

$$E(\boldsymbol{w}) \triangleq \sum_{\boldsymbol{r}\in\mathcal{R}} E_{\boldsymbol{r}}(\boldsymbol{w}), \tag{1}$$

where

$$E_{\boldsymbol{r}}(\boldsymbol{w}) \triangleq \alpha\,\Phi\big(A_{\boldsymbol{r}}(\boldsymbol{w})\big) + \beta\,\Psi\big(B_{\boldsymbol{r}}(\boldsymbol{w})\big). \tag{2}$$

Here, $\Phi(\xi)$ and $\Psi(\xi)$ are concave nondecreasing differentiable penalty functions on $[0,\infty)$, and $\alpha \geq 0$ and $\beta \geq 0$ determine the relative influence of the respective penalty terms. Furthermore,

$$A_{\boldsymbol{r}}(\boldsymbol{w}) \triangleq \big(D_{\boldsymbol{r}} + G_{\boldsymbol{r}}^{(x)} u_{\boldsymbol{r}} + G_{\boldsymbol{r}}^{(y)} v_{\boldsymbol{r}}\big)^2, \tag{3}$$

with

$$G_{\boldsymbol{r}}^{(x)} \triangleq \left.\frac{\partial G(\boldsymbol{x})}{\partial x}\right|_{\boldsymbol{x}=\boldsymbol{x}_{\boldsymbol{r}}+\bar{\boldsymbol{w}}_{\boldsymbol{r}}}, \quad G_{\boldsymbol{r}}^{(y)} \triangleq \left.\frac{\partial G(\boldsymbol{x})}{\partial y}\right|_{\boldsymbol{x}=\boldsymbol{x}_{\boldsymbol{r}}+\bar{\boldsymbol{w}}_{\boldsymbol{r}}} \tag{4}$$

and

$$D_{\boldsymbol{r}} \triangleq G(\boldsymbol{x}_{\boldsymbol{r}} + \bar{\boldsymbol{w}}_{\boldsymbol{r}}) - F(\boldsymbol{x}_{\boldsymbol{r}}) - G_{\boldsymbol{r}}^{(x)}\bar{u}_{\boldsymbol{r}} - G_{\boldsymbol{r}}^{(y)}\bar{v}_{\boldsymbol{r}}, \tag{5}$$

quantifies violation of the brightness constancy constraint [2]. We note that $A_{\boldsymbol{r}}(\boldsymbol{w})$ is obtained by squaring the linear approximation of

This work was supported in part by the Austrian Science Fund (FWF) under grant P 32055-N31.

the difference between $G(\boldsymbol{x_r}+\boldsymbol{w_r})$ and $F(\boldsymbol{x_r})$ around a given OF value $\bar{\boldsymbol{w}}$, where $\bar{\boldsymbol{w}}_{\boldsymbol{r}}=(\bar{u}_{\boldsymbol{r}}\ \bar{v}_{\boldsymbol{r}})^{\mathrm{T}}$ denotes the $\boldsymbol{r}$-th component vector of $\bar{\boldsymbol{w}}$. Thus

$$A_{\boldsymbol{r}}(\boldsymbol{w}) \approx \big(G(\boldsymbol{x_r}+\boldsymbol{w_r})-F(\boldsymbol{x_r})\big)^2. \tag{6}$$

Finally,

$$B_{\boldsymbol{r}}(\boldsymbol{w}) \triangleq \sum_{\boldsymbol{\delta}\in\mathcal{N}} \big((u_{\boldsymbol{r}+\boldsymbol{\delta}}-u_{\boldsymbol{r}})^2+(v_{\boldsymbol{r}+\boldsymbol{\delta}}-v_{\boldsymbol{r}})^2\big), \tag{7}$$

with $\mathcal{N}\triangleq\{(1,0),(0,1)\}$, quantifies violation of the spatial smoothness constraint [2].

We adopt a Bayesian perspective of OF estimation by considering $\boldsymbol{w}$, $F(\boldsymbol{x})$, and $G(\boldsymbol{x})$ as random elements and interpreting $E(\boldsymbol{w})$ in (1), (2) as the energy of a Markov random field [5], [10, Sec. 8.3]. This allows us to introduce a posterior pdf of $\boldsymbol{w}$ as

$$p(\boldsymbol{w}|F,G) \triangleq \frac{h(\boldsymbol{w})}{Z} \quad \text{with } h(\boldsymbol{w}) \triangleq \exp\big(-E(\boldsymbol{w})\big), \tag{8}$$

where $Z\triangleq\int\exp\big(-E(\boldsymbol{w})\big)\,\mathrm{d}\boldsymbol{w}<\infty$. Inserting (1) and (2) in (8), we obtain $h(\boldsymbol{w})=\mathcal{L}(\boldsymbol{w})\mathcal{P}(\boldsymbol{w})$ with

$$\mathcal{L}(\boldsymbol{w}) \triangleq \exp\bigg(-\alpha\sum_{\boldsymbol{r}\in\mathcal{R}}\Phi\big(A_{\boldsymbol{r}}(\boldsymbol{w})\big)\bigg), \tag{9}$$

$$\mathcal{P}(\boldsymbol{w}) \triangleq \exp\bigg(-\beta\sum_{\boldsymbol{r}\in\mathcal{R}}\Psi\big(B_{\boldsymbol{r}}(\boldsymbol{w})\big)\bigg). \tag{10}$$

Here, $\mathcal{L}(\boldsymbol{w})$ describes the statistical relation between $\boldsymbol{w}$ and the images F, G and can be interpreted as a likelihood function, and $\mathcal{P}(\boldsymbol{w})$ can be interpreted as an improper prior. Further technical details regarding our Bayesian perspective (including conditions for $Z<\infty$) are discussed in the supplementary material [11, Sec. S4].

3. VARIATIONAL LOWER BOUND

The proposed Bayesian methods for OF estimation are based on an upper bound on the energy $E(\boldsymbol{w})$ or, equivalently, a lower bound on $h(\boldsymbol{w})$. This bound is obtained by lower-bounding the functions $\Phi(\xi)$ and $\Psi(\xi)$. As shown in [12, Sec. 3.3], for any closed concave nondecreasing function $\Phi(\xi)$, there exists a unique closed concave *conjugate* defined as[1]

$$\Phi_\star(\lambda) \triangleq \inf_{\xi\geq 0}\{\lambda\xi-\Phi(\xi)\}. \tag{11}$$

Here, $\lambda\geq 0$ is referred to as the variational parameter. By (11), we have $\lambda\xi-\Phi(\xi)\geq\Phi_\star(\lambda)$ or equivalently $\Phi(\xi)\leq\lambda\xi-\Phi_\star(\lambda)$, for any $\lambda\geq 0$. The last inequality can be used to obtain the upper bounds $\Phi\big(A_{\boldsymbol{r}}(\boldsymbol{w})\big)\leq\lambda_{\boldsymbol{r}}A_{\boldsymbol{r}}(\boldsymbol{w})-\Phi_\star(\lambda_{\boldsymbol{r}})$ and, similarly, $\Psi\big(B_{\boldsymbol{r}}(\boldsymbol{w})\big)\leq\mu_{\boldsymbol{r}}B_{\boldsymbol{r}}(\boldsymbol{w})-\Psi_\star(\mu_{\boldsymbol{r}})$, for any $\lambda_{\boldsymbol{r}},\mu_{\boldsymbol{r}}\geq 0$. Inserting these bounds in (2) gives $E_{\boldsymbol{r}}(\boldsymbol{w})\leq E_{\boldsymbol{r}}(\boldsymbol{w};\boldsymbol{\lambda_r})$ with

$$E_{\boldsymbol{r}}(\boldsymbol{w};\boldsymbol{\lambda_r}) \triangleq \alpha\big(\lambda_{\boldsymbol{r}}A_{\boldsymbol{r}}(\boldsymbol{w})-\Phi_\star(\lambda_{\boldsymbol{r}})\big)+\beta\big(\mu_{\boldsymbol{r}}B_{\boldsymbol{r}}(\boldsymbol{w})-\Psi_\star(\mu_{\boldsymbol{r}})\big), \tag{12}$$

where $\boldsymbol{\lambda_r}\triangleq\big(\lambda_{\boldsymbol{r}}\ \ \mu_{\boldsymbol{r}}\big)^{\mathrm{T}}\in\mathbb{R}^2_+$. Using (1), we furthermore obtain

$$E(\boldsymbol{w})\leq E(\boldsymbol{w};\boldsymbol{\lambda}) \quad \text{with } E(\boldsymbol{w};\boldsymbol{\lambda})\triangleq\sum_{\boldsymbol{r}\in\mathcal{R}}E_{\boldsymbol{r}}(\boldsymbol{w};\boldsymbol{\lambda_r}), \tag{13}$$

where the vector $\boldsymbol{\lambda}$ of length $2NM$ stacks all the $\boldsymbol{\lambda_r}$. Finally, using (8), we obtain the *variational lower bound*

$$h(\boldsymbol{w})\geq h(\boldsymbol{w};\boldsymbol{\lambda}) \quad \text{with } h(\boldsymbol{w};\boldsymbol{\lambda})\triangleq\exp\big(-E(\boldsymbol{w};\boldsymbol{\lambda})\big). \tag{14}$$

[1] In (11), we implicitly follow the standard approach from convex analysis and assume that the convex function $-\Phi(\xi)$ and its (convex) conjugate $-\Phi_\star(\lambda)$ are proper closed extended-valued functions, i.e., with range in $\mathbb{R}\cup\{\infty\}$. Thus, the concave functions $\Phi(\xi)$ and $\Phi_\star(\lambda)$ have range $\mathbb{R}\cup\{-\infty\}$. In our application context, it suffices to restrict their domain to $\mathbb{R}_+$.

The "energy upper bound" $E(\boldsymbol{w};\boldsymbol{\lambda})$ in (13) is convex in $\boldsymbol{\lambda}$ and quadratic in $\boldsymbol{w}$. The first property can be shown as follows. According to (11), $\Phi_\star(\lambda)$ is defined as the pointwise minimum of a family of affine functions of λ, which implies that it is concave [12, Sec. 3.3], so $-\Phi^\star(\lambda)$ is convex. Therefore, we conclude from (12) that $E_{\boldsymbol{r}}(\boldsymbol{w};\boldsymbol{\lambda_r})$ is convex in $\boldsymbol{\lambda_r}$. Finally, since $E(\boldsymbol{w};\boldsymbol{\lambda})$ is the sum of convex functions, it is itself convex in $\boldsymbol{\lambda}$. Next, inserting (3) and (7) into (12) and the resulting expression into (13) yields

$$\begin{aligned}E(\boldsymbol{w};\boldsymbol{\lambda})=\sum_{\boldsymbol{r}\in\mathcal{R}}\Big(&\alpha\Big(\lambda_{\boldsymbol{r}}\big(D_{\boldsymbol{r}}+G^{(x)}_{\boldsymbol{r}}u_{\boldsymbol{r}}+G^{(y)}_{\boldsymbol{r}}v_{\boldsymbol{r}}\big)^2\\&-\Phi_\star(\lambda_{\boldsymbol{r}})\Big)+\beta\Big(\mu_{\boldsymbol{r}}\sum_{\boldsymbol{\delta}\in\mathcal{N}}\big((u_{\boldsymbol{r}+\boldsymbol{\delta}}-u_{\boldsymbol{r}})^2\\&+(v_{\boldsymbol{r}+\boldsymbol{\delta}}-v_{\boldsymbol{r}})^2\big)-\Psi_\star(\mu_{\boldsymbol{r}})\Big)\Big).\end{aligned} \tag{15}$$

This is seen to be a quadratic function of $u_{\boldsymbol{r}}$ and $v_{\boldsymbol{r}}$ for any $\boldsymbol{r}\in\mathcal{R}$. Thus, we can write $E(\boldsymbol{w};\boldsymbol{\lambda})$ up to an additive constant as a quadratic form $(\boldsymbol{w}-\boldsymbol{w}_0)^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{w}-\boldsymbol{w}_0)$ with some offset $\boldsymbol{w}_0$ and matrix $\boldsymbol{J}$. As we show in [11, Sec. S4], $\boldsymbol{J}$ is positive definite. This implies that

$$q(\boldsymbol{w};\boldsymbol{\lambda}) \triangleq \frac{h(\boldsymbol{w};\boldsymbol{\lambda})}{Z(\boldsymbol{\lambda})}=\frac{\exp\big(-E(\boldsymbol{w};\boldsymbol{\lambda})\big)}{Z(\boldsymbol{\lambda})}, \tag{16}$$

with $Z(\boldsymbol{\lambda})\triangleq\int\exp\big(-E(\boldsymbol{w};\boldsymbol{\lambda})\big)\,\mathrm{d}\boldsymbol{w}<\infty$, is a Gaussian pdf with mean $\boldsymbol{w}_0$ and precision matrix $\boldsymbol{J}$. As a consequence, $Z(\boldsymbol{\lambda})$ can be determined easily. Motivated by (8) and (14), we use $q(\boldsymbol{w};\boldsymbol{\lambda})$ as an approximation of the posterior pdf $p(\boldsymbol{w}|F,G)$.

4. BAYESIAN OF ESTIMATION METHODS

We now use the variational lower bound derived in the previous section to develop two Bayesian methods for OF estimation.

4.1. MAP Estimation

In the MAP framework, the OF estimate is ideally obtained by minimizing the energy function $E(\boldsymbol{w})$ in (1) or, equivalently, by maximizing $h(\boldsymbol{w})$ in (8). Note that this is indeed MAP estimation since, by (8), $h(\boldsymbol{w})$ equals the posterior pdf $p(\boldsymbol{w}|F,G)$ up to a normalization constant. Because maximizing $h(\boldsymbol{w})$ is difficult, we instead maximize the lower bound $h(\boldsymbol{w};\boldsymbol{\lambda})$ in (14) with respect to both $\boldsymbol{w}$ and $\boldsymbol{\lambda}$ in an iterative, alternating manner. Methods of this kind are called "Type I" methods in [6]. More precisely, we update the variational parameter estimate $\hat{\boldsymbol{\lambda}}$ and subsequently the OF estimate $\hat{\boldsymbol{w}}$ according to

$$\hat{\boldsymbol{\lambda}}=\operatorname*{arg\,max}_{\boldsymbol{\lambda}\succeq\boldsymbol{0}}\ \log h(\hat{\boldsymbol{w}};\boldsymbol{\lambda}), \tag{17}$$

$$\hat{\boldsymbol{w}}=\operatorname*{arg\,max}_{\boldsymbol{w}}\ \log h(\boldsymbol{w};\hat{\boldsymbol{\lambda}}), \tag{18}$$

where $\succeq$ denotes elementwise $\geq$. The recursion (17), (18) is repeated until a stopping criterion is satisfied.

As we show in [11, Sec. S1], the maximization (17) has a closed-form solution given by

$$\hat{\lambda}_{\boldsymbol{r}}=\Phi'\big(A_{\boldsymbol{r}}(\hat{\boldsymbol{w}})\big),\ \ \hat{\mu}_{\boldsymbol{r}}=\Psi'\big(B_{\boldsymbol{r}}(\hat{\boldsymbol{w}})\big),\ \ \boldsymbol{r}\in\mathcal{R}, \tag{19}$$

where the prime denotes the derivative. Furthermore, as also shown in [11, Sec. S1], the maximization (18) amounts to solving the system of $2NM$ linear equations

$$\frac{\partial}{\partial u_{\boldsymbol{r}}}E(\boldsymbol{w};\hat{\boldsymbol{\lambda}})=0,\ \ \frac{\partial}{\partial v_{\boldsymbol{r}}}E(\boldsymbol{w};\hat{\boldsymbol{\lambda}})=0,\ \ \boldsymbol{r}\in\mathcal{R} \tag{20}$$

for $\boldsymbol{w}$. Here, expressions of the partial derivatives $\frac{\partial}{\partial u_{\boldsymbol{r}}}E(\boldsymbol{w};\hat{\boldsymbol{\lambda}})$ and $\frac{\partial}{\partial v_{\boldsymbol{r}}}E(\boldsymbol{w};\hat{\boldsymbol{\lambda}})$ are provided in [11, Eqs. (S4), (S5)].

The recursion (17), (18) can be modified with the goal of reducing the error in the approximation (6), which tends to be large when $\|\boldsymbol{w}_{\boldsymbol{r}}-\bar{\boldsymbol{w}}_{\boldsymbol{r}}\|$ is large [2]. A commonly used strategy is to update the linearization point $\bar{\boldsymbol{w}}_{\boldsymbol{r}}$ in each iteration. The resulting modified algorithm is stated in [11, Sec. S3]. This algorithm is an instance of the half-quadratic minimization algorithm [13], which has been used, e.g., in [2] and [3]. However, our Bayesian formulation makes it possible to derive quantities characterizing the reliability (uncertainty) of the OF estimate, as will be discussed in Section 4.3.

4.2. VB Estimation

VB methods rely on the choice of an approximation of the posterior pdf from a family of tractable pdfs [14]. Here, we employ this strategy to find an optimal value of the variational parameter vector $\boldsymbol{\lambda}$. Following [6], we define this optimal value, denoted $\boldsymbol{\lambda}^*$, as the value of $\boldsymbol{\lambda}$ minimizing $\int \big|h(\boldsymbol{w}) - h(\boldsymbol{w};\boldsymbol{\lambda})\big| \,\mathrm{d}\boldsymbol{w}$. A connection to the VB principle is established by the fact that $h(\boldsymbol{w})$ and $h(\boldsymbol{w};\boldsymbol{\lambda})$ are (up to normalization) the posterior pdf and its approximation, respectively. Methods of this kind are called "Type II" methods in [6].

By (14), the absolute value in the integral $\int \big|h(\boldsymbol{w})-h(\boldsymbol{w};\boldsymbol{\lambda})\big| \,\mathrm{d}\boldsymbol{w}$ can be dropped. Thus, we obtain

$$\boldsymbol{\lambda}^* = \operatorname*{arg\,max}_{\boldsymbol{\lambda}\succeq\mathbf{0}} \int h(\boldsymbol{w};\boldsymbol{\lambda}) \,\mathrm{d}\boldsymbol{w}\,. \tag{21}$$

Following [15], we perform the maximization in (21) using the expectation maximization (EM) algorithm presented in [16]. Let

$$\begin{aligned} Q(\boldsymbol{\lambda},\hat{\boldsymbol{\lambda}}) &\triangleq \int h(\boldsymbol{w};\hat{\boldsymbol{\lambda}}) \,\log h(\boldsymbol{w};\boldsymbol{\lambda}) \,\mathrm{d}\boldsymbol{w} \\ &= Z(\hat{\boldsymbol{\lambda}}) \int q(\boldsymbol{w};\hat{\boldsymbol{\lambda}}) \log h(\boldsymbol{w};\boldsymbol{\lambda}) \,\mathrm{d}\boldsymbol{w} \\ &= Z(\hat{\boldsymbol{\lambda}})\, \mathrm{E}^{(\hat{\boldsymbol{\lambda}})}\big\{\log h(\boldsymbol{w};\boldsymbol{\lambda})\big\}, \end{aligned} \tag{22}$$

where (16) was used in the second line and $\mathrm{E}^{(\hat{\boldsymbol{\lambda}})}\{\cdot\}$ denotes expectation with respect to the pdf $q(\boldsymbol{w};\hat{\boldsymbol{\lambda}})$. The EM algorithm alternates between the E-step and the M-step until a stopping criterion is satisfied. In the proposed algorithm, in each iteration, we perform the M-step first because in our context it is easier to initialize. To obtain a tractable algorithm, we restrict the precision matrix $\boldsymbol{J}$ to a diagonal matrix (retaining the same diagonal elements).[2] With this simplification, under the distribution $q(\boldsymbol{w};\hat{\boldsymbol{\lambda}})$, $u_{\boldsymbol{r}}$ and $v_{\boldsymbol{r}'}$ are now uncorrelated for all $\boldsymbol{r},\boldsymbol{r}'\in\mathcal{R}$, and furthermore $u_{\boldsymbol{r}}$ and $u_{\boldsymbol{r}'}$ as well as $v_{\boldsymbol{r}}$ and $v_{\boldsymbol{r}'}$ are uncorrelated unless $\boldsymbol{r}=\boldsymbol{r}'$.

In the M-step, $\hat{\boldsymbol{\lambda}}$ is updated by maximizing $Q(\boldsymbol{\lambda},\hat{\boldsymbol{\lambda}})$ with respect to $\boldsymbol{\lambda}$, i.e.,

$$\hat{\boldsymbol{\lambda}} \leftarrow \operatorname*{arg\,max}_{\boldsymbol{\lambda}\succeq\mathbf{0}} Q(\boldsymbol{\lambda},\hat{\boldsymbol{\lambda}})\,. \tag{23}$$

As we show in [11, Sec. S2], this amounts to

$$\hat{\lambda}_{\boldsymbol{r}} \leftarrow \Phi'\big(\mathrm{E}^{(\hat{\boldsymbol{\lambda}})}\big\{A_{\boldsymbol{r}}(\boldsymbol{w})\big\}\big)\,,\quad \hat{\mu}_{\boldsymbol{r}} \leftarrow \Psi'\big(\mathrm{E}^{(\hat{\boldsymbol{\lambda}})}\big\{B_{\boldsymbol{r}}(\boldsymbol{w})\big\}\big)\,, \tag{24}$$

for all $\boldsymbol{r}\in\mathcal{R}$. Closed form expressions for $\mathrm{E}^{(\hat{\boldsymbol{\lambda}})}\big\{A_{\boldsymbol{r}}(\boldsymbol{w})\big\}$ and $\mathrm{E}^{(\hat{\boldsymbol{\lambda}})}\big\{B_{\boldsymbol{r}}(\boldsymbol{w})\big\}$ that can be evaluated in terms of the posterior means $\hat{u}_{\boldsymbol{r}}$ and $\hat{v}_{\boldsymbol{r}}$ and posterior variances $s_{\boldsymbol{r}}^{(u)}$ and $s_{\boldsymbol{r}}^{(v)}$ of $u_{\boldsymbol{r}}$ and $v_{\boldsymbol{r}}$ are provided in [11, Eqs. (S8), (S9)].

In the E-step, the posterior means $\hat{\boldsymbol{w}}_{\boldsymbol{r}} = (\hat{u}_{\boldsymbol{r}}\ \hat{v}_{\boldsymbol{r}})^{\mathrm{T}}$ and posterior variances $s_{\boldsymbol{r}}^{(u)}$ and $s_{\boldsymbol{r}}^{(v)}$ are updated for all $\boldsymbol{r}\in\mathcal{R}$. Because $q(\boldsymbol{w};\hat{\boldsymbol{\lambda}})$ is a Gaussian pdf, its mean equals its mode, which was calculated by the MAP method (see (18)). That is, the new posterior means are the solutions of the system of linear equations (20). Finally, with knowledge of $\hat{\lambda}_{\boldsymbol{r}}$ and $\hat{\mu}_{\boldsymbol{r}}$ for $\boldsymbol{r}\in\mathcal{R}$, the posterior variances can be evaluated in closed form as stated in [11, Eqs. (S11), (S12)].

As in Section 4.1, we update the linearization point $\bar{\boldsymbol{w}}_{\boldsymbol{r}}$ in each iteration to reduce the approximation error in (6). The resulting algorithm is stated in [11, Sec. S3]. This algorithm generalizes the algorithms presented in [4, 5] in that it allows for the use of a wider range of likelihood functions and prior pdfs.

4.3. Discussion

Upon convergence, the VB method naturally produces an approximation of the posterior pdf $p(\boldsymbol{w}|F,G)$, namely, $q(\boldsymbol{w};\hat{\boldsymbol{\lambda}})$ in (16), where $\hat{\boldsymbol{\lambda}}$ is the final value of the variational parameter returned after termination. Interestingly, this is true also for the MAP method, a fact that was not noted in [2], [3]. Thus, our variational framework provides a unified view of the VB and MAP methods in the sense that both calculate a (Gaussian) approximate posterior pdf $q(\boldsymbol{w};\hat{\boldsymbol{\lambda}})$. In both methods, the OF estimate $\hat{\boldsymbol{w}}$ is given by the mode of $q(\boldsymbol{w};\hat{\boldsymbol{\lambda}})$, and the uncertainty of the estimate can be characterized by the entropy $h_{\boldsymbol{r}} \triangleq -\mathrm{E}^{(\hat{\boldsymbol{\lambda}})}\big\{\log q(\boldsymbol{w}_{\boldsymbol{r}};\hat{\boldsymbol{\lambda}})\big\}$ [5]. By the Gaussianity of $q(\boldsymbol{w};\hat{\boldsymbol{\lambda}})$ and our assumption that $u_{\boldsymbol{r}}$ and $v_{\boldsymbol{r}}$ are uncorrelated under this distribution, the entropy is given by [5]

$$h_{\boldsymbol{r}} = \frac{1}{2}\big(\log s_{\boldsymbol{r}}^{(u)} + \log s_{\boldsymbol{r}}^{(v)}\big) + c\,, \tag{25}$$

where c is an irrelevant constant. Here, both methods calculate the variances $s_{\boldsymbol{r}}^{(u)}$ and $s_{\boldsymbol{r}}^{(v)}$ according to the closed form expressions given in [11, Eqs. (S11), (S12)], using the final values of $\hat{\lambda}_{\boldsymbol{r}}$ and $\hat{\mu}_{\boldsymbol{r}}$ after termination. Note that within this unified view of the VB and MAP methods, the two methods differ merely in the way the variational parameter estimate $\hat{\boldsymbol{\lambda}}$ is calculated.

5. DOMAIN-SPECIFIC LIKELIHOOD FUNCTIONS

A major advantage of the proposed Bayesian methods, besides providing an approximate posterior pdf and not just a point estimate, is that they admit a wide range of domain-specific likelihood functions $\mathcal{L}(\boldsymbol{w})$ and priors $\mathcal{P}(\boldsymbol{w})$. This flexibility is due to the formulation of $\mathcal{L}(\boldsymbol{w})$ and $\mathcal{P}(\boldsymbol{w})$ in terms of the penalty functions $\Phi(\xi)$ and $\Psi(\xi)$ according to (9) and (10), combined with the fact that $\Phi(\xi)$ and $\Psi(\xi)$ are merely required to be concave and nondecreasing.

Here, we focus on likelihood functions designed for the ultrasound imaging modality [19]. Concretely, we consider the likelihood functions proposed in [8] and [9], which we refer to as $\mathcal{L}_{\mathrm{CD}_2}(\boldsymbol{w})$ and $\mathcal{L}_{\mathrm{MS}}(\boldsymbol{w})$, respectively. Both admit the expression (9) except that in [8] and [9], $A_{\boldsymbol{r}}(\boldsymbol{w})$ in (3) is replaced by $\big(G(\boldsymbol{x}_{\boldsymbol{r}}+\boldsymbol{w}_{\boldsymbol{r}}) - F(\boldsymbol{x}_{\boldsymbol{r}})\big)^2$ (see (6)). We will use $A_{\boldsymbol{r}}(\boldsymbol{w})$ instead, which provides a good approximation as long as $\|\boldsymbol{w}_{\boldsymbol{r}} - \bar{\boldsymbol{w}}_{\boldsymbol{r}}\|$ is small. The corresponding penalty functions, to be used in (9), are

$$\begin{aligned} \Phi_{\mathrm{CD}_2}(\xi) &\triangleq 2\log\big(1+\exp(\xi')\big) - \xi', \\ \Phi_{\mathrm{MS}}(\xi) &\triangleq \Phi_{\mathrm{CD}_2}(\xi) + \frac{3}{2}\log\left(1 - \frac{4\rho\exp(\xi')}{(\exp(\xi')+1)^2}\right). \end{aligned}$$

Here, $\xi' \triangleq \frac{2\sqrt{\xi}}{b}$, where $b>0$ is a log-compression factor (i.e., the B-mode [19, Ch. 10] images $F(\boldsymbol{x})$ and $G(\boldsymbol{x})$ are obtained by applying the $b\log(\cdot)$ transformation to the corresponding envelopes), and $\rho\in[0,1)$ is the noise correlation coefficient of $F(\boldsymbol{x})$ and $G(\boldsymbol{x})$ (see [9] for details). Finally, we consider a third likelihood function,

[2] This restriction is equivalent to replacing $q(\boldsymbol{w};\boldsymbol{\lambda})$ by the nearest (in terms of minimal Kullback-Leibler divergence) fully factorizing pdf [17]. In this view, our algorithm can be interpreted as a variational EM algorithm using the mean-field approximation [18].

denoted as $\mathcal{L}_{\mathrm{C}}(\boldsymbol{w})$, that uses the penalty function $\Phi_{\mathrm{C}}(\xi) \triangleq \sqrt{\xi+\epsilon}$ with a small $\epsilon > 0$ [2]. This likelihood function is not related to the ultrasound modality. All three penalty functions $\Phi_{\mathrm{CD}_2}(\xi)$, $\Phi_{\mathrm{MS}}(\xi)$, and $\Phi_{\mathrm{C}}(\xi)$ are concave and nondecreasing.

6. NUMERICAL RESULTS

We present numerical results[3] for the synthetic cardiac ultrasound dataset STRAUS [20]. From each 3D volume in that dataset, we extracted a slice to obtain 2D image data. We compare the performance obtained with the proposed MAP and VB methods using the likelihood functions $\mathcal{L}_{\mathrm{CD}_2}(\boldsymbol{w})$, $\mathcal{L}_{\mathrm{MS}}(\boldsymbol{w})$, and $\mathcal{L}_{\mathrm{C}}(\boldsymbol{w})$. For $\Psi(\xi)$ (involved in the prior $\mathcal{P}(\boldsymbol{w})$, see (10)), we use the penalty function $\sqrt{\xi+\epsilon}$ (equal to $\Phi_{\mathrm{C}}(\xi)$). We consider the following performance metrics: the average endpoint error (AEE) [1] calculated by averaging the endpoint error over all pixels of all sequences; the area under the sparsification plot (AUS) [5], and the "conditional likelihood" (CL) calculated by evaluating the logarithm of the approximate posterior at the ground-truth value of the OF [21]. We note that the highest CL is achieved when the mean of the approximate posterior coincides with the ground truth and the variance is minimal. For comparison, we also present the AEE for the state-of-the-art methods proposed in [22] and [23], which we refer to as "SPtrack" and "Monogenic," respectively. Since these methods only provide point estimates of the OF, the AUS and CL metrics cannot be computed. The parameters of all methods were optimized using the *ladprox* sequence of the STRAUS dataset. This was done by minimizing the CL metric in the case of the VB method and the AEE metric in the case of the MAP method and the benchmark methods [22, 23]. We note that for the MAP method, minimizing the CL metric did not yield the best overall performance.

An example of the results of the VB method using the likelihood function $\mathcal{L}_{\mathrm{C}}(\boldsymbol{w})$ is shown in Fig. 1. Note that the high endpoint error at the bottom left part of the heart muscle (Fig. 1(c)) is matched by a high uncertainty (Fig. 1(d)). Further numerical results are provided in Table 1 for data corresponding to short-axis and long-axis scans of the heart [20]. The reported metrics are based on measuring the OF in units of pixels (px); the pixel size is $0.67{\times}0.85\,\mathrm{mm}^2$ for the short-axis scan and $0.67 \times 0.58\,\mathrm{mm}^2$ for the long-axis scan. We see that both the MAP and VB methods yield significantly smaller AEE values than the two benchmark methods [22, 23]. Furthermore, for the same likelihood function, the VB method consistently outperforms the MAP method: it is more accurate (indicated by a moderately smaller AEE), yields more faithful uncertainty estimates (indicated by a significantly smaller AUS), and its approximate posterior has a larger CL. Therefore, we further consider only the VB method.

[3] Source code and data used to generate the numerical results are available at https://github.com/deu439/BOF.

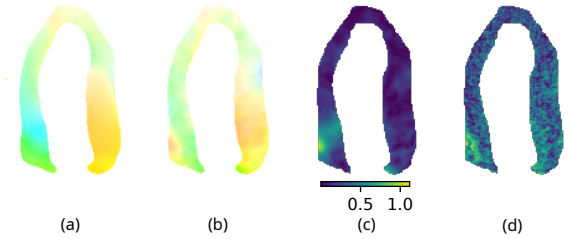

Fig. 1. Example result of the VB method with likelihood function $\mathcal{L}_{\mathrm{C}}(\boldsymbol{w})$ for the *ladprox* sequence; long-axis scan. (a) True OF, (b) OF estimate, (c) endpoint error (in [px], with color bar), (d) uncertainty (entropy, see (25)). In (a) and (b), hue and saturation indicate OF direction and magnitude, respectively.

Regarding the influence of the likelihood function on the VB method, we see that the ultrasound-specific likelihood function $\mathcal{L}_{\mathrm{MS}}(\boldsymbol{w})$ leads to a lower AEE and a higher CL compared to $\mathcal{L}_{\mathrm{C}}(\boldsymbol{w})$, but the AUS is larger. Furthermore, $\mathcal{L}_{\mathrm{CD}_2}(\boldsymbol{w})$ consistently yields poorer results than $\mathcal{L}_{\mathrm{MS}}(\boldsymbol{w})$. We conjecture that this is because $\mathcal{L}_{\mathrm{CD}_2}(\boldsymbol{w})$, contrary to $\mathcal{L}_{\mathrm{MS}}(\boldsymbol{w})$, does not take into account the noise correlation of $F(\boldsymbol{x})$ and $G(\boldsymbol{x})$, which is significant for the used dataset [20]. Overall, the VB method using the ultrasound-specific likelihood function $\mathcal{L}_{\mathrm{MS}}(\boldsymbol{w})$ achieves the best performance, except that $\mathcal{L}_{\mathrm{C}}(\boldsymbol{w})$ leads to better uncertainty estimates.

Finally, the runtime of our nonoptimized single-threaded Python 3.10 implementation of the VB method on an Intel(R) Core(TM) i5-7500 CPU with 3.40 GHz clock rate was around 1.8 s per image pair, for image size $208{\times}224$ px.

7. CONCLUSION

We developed a Bayesian framework for OF estimation that uses a variational lower bound to obtain a variational approximation of the posterior distribution. Our framework provides a unified variational formulation of the MAP and VB approaches to Bayesian OF estimation. Moreover, it is able to accommodate domain-specific likelihood functions and prior distributions, and it provides a quantitative measure of the uncertainty of OF estimation. Numerical experiments on synthetic cardiac ultrasound data confirmed the importance of these features. In particular, we observed that the use of ultrasound-specific likelihood functions improves the performance of Bayesian OF estimation, and that the VB method outperforms the MAP method. Interesting directions for future research include the optimal selection of the penalty function defining the prior distribution and the use of locally-adaptive likelihood functions such as [24], [25] within our Bayesian OF estimation framework.

	Short-axis Scan			***Long-axis Scan***		
	AEE [px]	AUS	CL	AEE [px]	AUS	CL
SPtrack [22]	0.462 (+32.9%)	—	—	0.421 (+47.0%)	—	—
Monogenic [23]	0.433 (+24.7%)	—	—	0.388 (+35.4%)	—	—
MAP (C)	0.348 (0.0%)	1.060 (0.0%)	$-4.58 \cdot 10^5$	0.287 (0.0%)	0.932 (0.0%)	$-6.43 \cdot 10^5$
MAP (CD_2)	0.370 (+6.3%)	1.036 (−4.7%)	$-6.53 \cdot 10^4$	0.291 (+1.5%)	0.888 (−4.7%)	$-5.51 \cdot 10^4$
MAP (MS)	0.340 (−2.2%)	1.077 (+1.6%)	$-6.81 \cdot 10^5$	0.271 (−5.6%)	1.054 (+13.1%)	$-1.02 \cdot 10^6$
VB (C)	0.345 (−0.7%)	**0.838 (−21.0%)**	$1.27 \cdot 10^4$	0.280 (−2.4%)	**0.681 (−26.9%)**	$1.20 \cdot 10^4$
VB (CD_2)	0.354 (+1.8%)	0.928 (−12.5%)	$1.23 \cdot 10^4$	0.289 (+0.8%)	0.748 (−19.7%)	$1.90 \cdot 10^4$
VB (MS)	**0.330 (−5.2%)**	0.889 (−16.2%)	$\mathbf{1.35 \cdot 10^4}$	**0.266 (−7.2%)**	0.718 (−23.0%)	$\mathbf{1.96 \cdot 10^4}$

Table 1. Performance of the MAP and VB methods using the likelihood functions $\mathcal{L}_{\mathrm{CD}_2}(\boldsymbol{w})$, $\mathcal{L}_{\mathrm{MS}}(\boldsymbol{w})$, and $\mathcal{L}_{\mathrm{C}}(\boldsymbol{w})$, obtained for STRAUS data corresponding to a short-axis scan (left) and a long-axis scan (right) of the heart. The % values in parentheses are the relative improvement with respect to the MAP method using $\mathcal{L}_{\mathrm{C}}(\boldsymbol{w})$, which is effectively equivalent to the benchmark method in [2].

8. REFERENCES

[1] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, Nov. 2010.

[2] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow based on a theory for warping," in *Proc. Eur. Conf. Comput. Vis.*, Prague, Czech Republic, May 2004, pp. 25–36.

[3] D. Sun, S. Roth, J. P. Lewis, and M. J. Black, "Learning optical flow," in *Proc. Eur. Conf. Comput. Vis.*, Marseille, France, Oct. 2008, pp. 83–97.

[4] G. Chantas, T. Gkamas, and C. Nikou, "Variational-Bayes optical flow," *J. Math. Imaging Vis.*, vol. 50, no. 3, pp. 199–214, Nov. 2014.

[5] A. S. Wannenwetsch, M. Keuper, and S. Roth, "Probflow: Joint optical flow and uncertainty estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, Dec. 2017, pp. 1182–1191.

[6] D. P. Wipf and B. D. Rao, "Latent variable Bayesian models for promoting sparsity," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6236–6255, Sept. 2011.

[7] Y. Marnissi, Y. Zheng, E. Chouzenoux, and J.-C. Pesquet, "A variational Bayesian approach for image restoration—Application to image deblurring with Poisson-Gaussian noise," *IEEE Trans. Comput. Imaging*, vol. 3, no. 4, pp. 722–737, May 2017.

[8] B. Cohen and I. Dinstein, "New maximum likelihood motion estimation schemes for noisy ultrasound images," *Pattern Recognit.*, vol. 35, no. 2, pp. 455–463, Feb. 2002.

[9] A. Myronenko, X. Song, and D. J. Sahn, "Maximum likelihood motion estimation in 3D echocardiography through non-rigid registration in spherical coordinates," in *Proc. Funct. Imaging Model. Hear.*, Nice, France, June 2009, pp. 427–436.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.

[11] J. Dorazil, B. H. Fleury, and F. Hlawatsch, "Bayesian methods for optical flow estimation using a variational approximation, with applications to ultrasound: Supplementary material," online: https://owncloud.tuwien.ac.at/index.php/s/WBPtLIUvZOJOpat, Mar. 2023.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[13] M. Nikolova and R. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1623–1627, May 2007.

[14] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Am. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, June 2017.

[15] J. A. Palmer, D. P. Wipf, K. Kreutz-Delgado, and B. Rao, "Variational EM algorithms for non-Gaussian latent variable models," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, Canada, Dec. 2005, vol. 18, pp. 1559–1566.

[16] J. Dauwels, S. Korl, and H. A. Loeliger, "Expectation maximization as message passing," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, SA, Australia, Sept. 2005, pp. 583–586.

[17] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders, "Variational Bayesian image restoration based on a product of t-distributions image prior," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1795–1805, Oct. 2008.

[18] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.

[19] T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*, Academic Press, Oxford, UK, second ed., 2014.

[20] M. Alessandrini et al., "A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: Methodology and open-access database," *IEEE Trans. Med. Imaging*, vol. 34, no. 7, pp. 1436–1451, Jan. 2015.

[21] J. Gast and S. Roth, "Lightweight probabilistic deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, June 2018, pp. 3369–3378.

[22] P. Vincent and D. Garcia, "Back to basics in ultrasound velocimetry: Tracking speckles by using a standard PIV algorithm," in *Proc. IEEE Int. Ultrason. Symp.*, Kobe, Japan, Oct. 2018.

[23] M. Alessandrini, A. Basarab, H. Liebgott, and O. Bernard, "Myocardial motion estimation from medical images using the monogenic signal," *IEEE Trans. Imag. Proces.*, vol. 22, no. 3, pp. 1084–1095, Nov. 2012.

[24] C. Wachinger, T. Klein, and N. Navab, "Locally adaptive Nakagami-based ultrasound similarity measures," *Ultrasonics*, vol. 52, no. 4, pp. 547–554, Apr. 2012.

[25] N. Ouzir, E. Ollila, and S. A. Vorobyov, "Data-adaptive similarity measures for B-mode ultrasound images using robust noise models," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 6, pp. 1244–1254, June 2020.