ProbFlow: Joint Optical Flow and Uncertainty Estimation
– Supplemental Material –

Anne S. Wannenwetsch[1] Margret Keuper[2] Stefan Roth[1]
[1]TU Darmstadt [2]University of Mannheim

Preface. In this supplemental material we derive update equations for the mean-field inference in Eq. (13) and show a proof for the solution of the Bayesian risk minimization in Eq. (8). We give further implementation details of ProbClassicA and ProbFlowFields, and present an analysis considering different design choices. Finally, we evaluate the performance of additional uncertainty measures and apply, for completeness, ProbFlowFields on the Middlebury benchmark [2].

A. Mean-field Update Equations

In the following, we show how to derive the mean-field update equations for ProbClassicA. Update equations for ProbFlowFields can be obtained similarly.

Notation. Note that strictly speaking the latent variables are given as $\mathbf{h} = (\mathbf{h}_{\gamma,ij,l})_{\gamma,ij}$ with $\gamma \in \{\mathrm{D}, \mathrm{S}_1, \dots, \mathrm{S}_p, \mathrm{N}_1, \dots, \mathrm{N}_q\}$, $p = |S(i,j)|$, $q = |N(i,j)|$ as we have separate latent variables for all penalty functions. In the following, we therefore use the notation $\mathbf{h}_{S_e}$, $\mathbf{h}_{N_e}$ and $f_{S_e}(\cdot)$, $f_{N_e}(\cdot)$ for $e \in S(i,j)$ and $e \in N(i,j)$, respectively. The flow vector of the corresponding neighboring pixel is denoted as $\mathbf{y}_e$. Moreover, GSM parameters π_l and σ_l differ for data, smoothness, and non-local potentials. For better readability, we drop indices D, S, and N that explicitly distinguish between the different GSMs.

Variational objective. As shown in Eq. (13), variational parameters $\boldsymbol{\theta}^\star$ are determined to minimize the Kullback-Leibler divergence between posterior p and its approximating distribution q, *i.e.*

$$\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}} D_{KL}\big(q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta}) \mid p(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h} \mid I)\big) \tag{22a}$$

$$= \arg\min_{\boldsymbol{\theta}} \underbrace{\mathbb{E}_{q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})}\big[\log q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})\big]}_{\text{①, entropy term}} - \underbrace{\mathbb{E}_{q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})}\big[\log p(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h} \mid I)\big]}_{②}. \tag{22b}$$

Recall that we defined the variational distribution q in Eq. (11) as

$$q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta}) = \prod_{i,j} \Big[q\left(\mathbf{y}_{ij}; \boldsymbol{\theta}\right) \cdot q\left(\hat{\mathbf{y}}_{ij}; \boldsymbol{\theta}\right) \cdot \prod_{\gamma} q\left(\mathbf{h}_{\gamma,ij}; \boldsymbol{\theta}\right) \Big]. \tag{23}$$

Then, the entropy term in ① can be split up as follows:

$$\mathbb{E}_{q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})}\Big[\log q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})\Big] = \mathbb{E}_{q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})}\left[\sum_{i,j} \log q\left(\mathbf{y}_{ij}; \boldsymbol{\theta}\right) + \sum_{i,j} \log q\left(\hat{\mathbf{y}}_{ij}; \boldsymbol{\theta}\right) + \sum_{i,j}\sum_{\gamma} \log q\left(\mathbf{h}_{\gamma,ij}; \boldsymbol{\theta}\right) \right] \tag{24a}$$

$$= \sum_{i,j} \mathbb{E}_{q(\mathbf{y}_{ij}; \boldsymbol{\theta})} \log q\left(\mathbf{y}_{ij}; \boldsymbol{\theta}\right) + \sum_{i,j} \mathbb{E}_{q(\hat{\mathbf{y}}_{ij}; \boldsymbol{\theta})} \log q\left(\hat{\mathbf{y}}_{ij}; \boldsymbol{\theta}\right) + \sum_{i,j}\sum_{\gamma} \mathbb{E}_{q(\mathbf{h}_{\gamma,ij}; \boldsymbol{\theta})} \log q\left(\mathbf{h}_{\gamma,ij}; \boldsymbol{\theta}\right). \tag{24b}$$

Using the well-known entropy of a Gaussian and a multinomial distribution, we obtain

$$\mathbb{E}_{q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})}\Big[\log q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})\Big] = -\frac{1}{2}\sum_{i,j} \log\big(\det(\boldsymbol{\Sigma}_{ij})\big) - \frac{1}{2}\sum_{i,j} \log\big(\det(\hat{\boldsymbol{\Sigma}}_{ij})\big) + \sum_{i,j}\sum_{\gamma}\sum_{l} k_{\gamma,ij,l}\ \log\ k_{\gamma,ij,l} + \text{const.} \tag{25}$$

In order to evaluate the term ② in Eq. (22b), it is necessary to compute

$$\begin{aligned}
&\log p(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h} \mid I) \\
&= \log \Bigg(\frac{1}{Z} \prod_{i,j} \bigg[\prod_l \pi_l^{h_{\mathrm{D},ij,l}} \mathcal{N}\Big(f_{\mathrm{D}}\big(\mathbf{y}_{ij}; I\big); 0, \sigma_l^2\Big)^{h_{\mathrm{D},ij,l}} \bigg]^{\lambda_{\mathrm{D}}} \cdot \prod_{e \in S(i,j)} \bigg[\prod_l \pi_l^{h_{\mathrm{S}_e,ij,l}} \mathcal{N}\Big(f_{\mathrm{S}_e}\big(\mathbf{y}_{ij}, \mathbf{y}_e\big); 0, \sigma_l^2\Big)^{h_{\mathrm{S}_e,ij,l}} \bigg]^{\lambda_{\mathrm{S}}} \\
&\qquad \cdot \exp\Big[- f_{\mathrm{C}}\big(\mathbf{y}_{ij}, \hat{\mathbf{y}}_{ij}\big)^2 \Big]^{\lambda_{\mathrm{C}}} \cdot \prod_{e \in N(i,j)} \bigg[\prod_l \pi_l^{h_{\mathrm{N}_e,ij,l}} \mathcal{N}\Big(f_{\mathrm{N}_e}\big(\hat{\mathbf{y}}_{ij}, \hat{\mathbf{y}}_e\big); 0, \sigma_l^2\Big)^{h_{\mathrm{N}_e,ij,l}} \bigg]^{\lambda_{\mathrm{N}}} \Bigg) && \text{(26a)} \\
&= \sum_{i,j} \Bigg[\lambda_{\mathrm{D}} \sum_l h_{\mathrm{D},ij,l} \Big(\log \pi_l + \log \mathcal{N}\Big(f_{\mathrm{D}}\big(\mathbf{y}_{ij}; I\big); 0, \sigma_l^2\Big)\Big) + \lambda_{\mathrm{S}} \sum_{e \in S(i,j)} \sum_l h_{\mathrm{S}_e,ij,l} \Big(\log \pi_l + \log \mathcal{N}\Big(f_{\mathrm{S}_e}\big(\mathbf{y}_{ij}, \mathbf{y}_e\big); 0, \sigma_l^2\Big)\Big) \\
&\qquad - \lambda_C \; f_{\mathrm{C}}\big(\mathbf{y}_{ij}, \hat{\mathbf{y}}_{ij}\big)^2 + \lambda_{\mathrm{N}} \sum_{e \in N(i,j)} \sum_l h_{\mathrm{N}_e,ij,l} \Big(\log \pi_l + \log \mathcal{N}\Big(f_{\mathrm{N}_e}\big(\hat{\mathbf{y}}_{ij}, \hat{\mathbf{y}}_e\big); 0, \sigma_l^2\Big)\Big) \Bigg] - \log Z && \text{(26b)} \\
&= \sum_{i,j} \Bigg[\lambda_{\mathrm{D}} \sum_l h_{\mathrm{D},ij,l} \bigg(\log \pi_l - \log \sigma_l - \frac{f_{\mathrm{D}}\big(\mathbf{y}_{ij}; I\big)^2}{2\sigma_l^2} \bigg) + \lambda_{\mathrm{S}} \sum_{e \in S(i,j)} \sum_l h_{\mathrm{S}_e,ij,l} \bigg(\log \pi_l - \log \sigma_l - \frac{f_{\mathrm{S}_e}\big(\mathbf{y}_{ij}, \mathbf{y}_e\big)^2}{2\sigma_l^2} \bigg) \\
&\qquad - \lambda_{\mathrm{C}} \; f_{\mathrm{C}}\big(\mathbf{y}_{ij}, \hat{\mathbf{y}}_{ij}\big)^2 + \lambda_{\mathrm{N}} \sum_{e \in N(i,j)} \sum_l h_{\mathrm{N}_e,ij,l} \bigg(\log \pi_l - \log \sigma_l - \frac{f_{\mathrm{N}_e}\big(\hat{\mathbf{y}}_{ij}, \hat{\mathbf{y}}_e\big)^2}{2\sigma_l^2} \bigg) \Bigg] + \text{const}, && \text{(26c)}
\end{aligned}$$

where we have defined $f_{\mathrm{C}}\big(\mathbf{y}_{ij}, \hat{\mathbf{y}}_{ij}\big) = \|\mathbf{y}_{ij} - \hat{\mathbf{y}}_{ij}\|_2$. We now take the expectation over $\mathbf{h}$ and simplify the remaining expectations as

$$\begin{aligned}
&\mathbb{E}_{q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})} \log p(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h} \mid I) \\
&= \mathbb{E}_{q(\mathbf{y}, \hat{\mathbf{y}}; \boldsymbol{\theta})} \sum_{i,j} \Bigg[\lambda_{\mathrm{D}} \sum_l k_{\mathrm{D},ij,l} \bigg(\log \pi_l - \log \sigma_l - \frac{f_{\mathrm{D}}(\mathbf{y}_{ij}; I)^2}{2\sigma_l^2} \bigg) \\
&\qquad + \lambda_{\mathrm{S}} \sum_{e \in S(i,j)} \sum_l k_{\mathrm{S}_e,ij,l} \bigg(\log \pi_l - \log \sigma_l - \frac{f_{\mathrm{S}_e}\big(\mathbf{y}_{ij}, \mathbf{y}_e\big)^2}{2\sigma_l^2} \bigg) \\
&\qquad - \lambda_{\mathrm{C}} \; f_{\mathrm{C}}\big(\mathbf{y}_{ij}, \hat{\mathbf{y}}_{ij}\big)^2 + \lambda_{\mathrm{N}} \sum_{e \in N(i,j)} \sum_l k_{\mathrm{N}_e,ij,l} \bigg(\log \pi_l - \log \sigma_l - \frac{f_{\mathrm{N}_e}\big(\hat{\mathbf{y}}_{ij}, \hat{\mathbf{y}}_e\big)^2}{2\sigma_l^2} \bigg) \Bigg] + \text{const} && \text{(27a)} \\
&= \sum_{i,j} \Bigg[\lambda_{\mathrm{D}} \sum_l k_{\mathrm{D},ij,l} \Big(\log \pi_l - \log \sigma_l \Big) + \lambda_{\mathrm{S}} \sum_{e \in S(i,j)} \sum_l k_{\mathrm{S}_e,ij,l} \Big(\log \pi_l - \log \sigma_l \Big) \\
&\qquad + \lambda_{\mathrm{N}} \sum_{e \in N(i,j)} \sum_l k_{\mathrm{N}_e,ij,l} \Big(\log \pi_l - \log \sigma_l \Big) \Bigg] \\
&\quad - \sum_{i,j} \Bigg[\lambda_{\mathrm{D}} \sum_l \frac{k_{\mathrm{D},ij,l}}{2\sigma_l^2} \underbrace{\mathbb{E}_{q(\mathbf{y}; \boldsymbol{\theta})} \; f_{\mathrm{D}}\big(\mathbf{y}_{ij}; I\big)^2}_{③,\, g_{\mathrm{D}}} + \lambda_{\mathrm{S}} \sum_{e \in S(i,j)} \sum_l \frac{k_{\mathrm{S}_e,ij,l}}{2\sigma_l^2} \underbrace{\mathbb{E}_{q(\mathbf{y}; \boldsymbol{\theta})} \; f_{\mathrm{S}_e}\big(\mathbf{y}_{ij}, \mathbf{y}_e\big)^2}_{④,\, g_{\mathrm{S}_e}} \\
&\qquad + \lambda_{\mathrm{C}} \underbrace{\mathbb{E}_{q(\mathbf{y}, \hat{\mathbf{y}}; \boldsymbol{\theta})} \; f_{\mathrm{C}}\big(\mathbf{y}_{ij}, \hat{\mathbf{y}}_{ij}\big)^2}_{⑤,\, g_{\mathrm{C}}} + \lambda_{\mathrm{N}} \sum_{e \in N(i,j)} \sum_l \frac{k_{\mathrm{N}_e,ij,l}}{2\sigma_l^2} \underbrace{\mathbb{E}_{q(\hat{\mathbf{y}}; \boldsymbol{\theta})} \; f_{\mathrm{N}_e}\big(\hat{\mathbf{y}}_{ij}, \hat{\mathbf{y}}_e\big)^2}_{⑥,\, g_{\mathrm{N}_e}} \Bigg] + \text{const}. && \text{(27b)}
\end{aligned}$$

To solve the expectation value w.r.t. the linearized brightness constancy in ③, we define $a = I_2\left(i + u_{ij}^0, j + v_{ij}^0\right) - I_1\left(i, j\right)$ and $\mathbf{b} = \nabla_2 I_2\left(i + u_{ij}^0, j + v_{ij}^0\right)^{\mathrm{T}}$. Then we have that

$$\begin{aligned} g_{\mathrm{D}}\big(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}; I\big) &= \mathbb{E}_{q(\mathbf{y};\boldsymbol{\theta})}\ f_{\mathrm{D}}\big(\mathbf{y}_{ij}; I\big)^2 && \text{(28a)} \\ &= \mathbb{E}_{q(\mathbf{y}_{ij};\boldsymbol{\theta})}\left[\Big(a + \mathbf{b}^{\mathrm{T}}\big(\mathbf{y}_{ij} - \mathbf{y}_{ij}^0\big)\Big)^2\right] && \text{(28b)} \\ &= \mathbb{E}_{q(\mathbf{y}_{ij};\boldsymbol{\theta})}\left[a^2 + 2a\mathbf{b}^{\mathrm{T}}\big(\mathbf{y}_{ij} - \mathbf{y}_{ij}^0\big) + \big(\mathbf{y}_{ij} - \mathbf{y}_{ij}^0\big)^{\mathrm{T}}\big(\mathbf{b}\mathbf{b}^{\mathrm{T}}\big)\big(\mathbf{y}_{ij} - \mathbf{y}_{ij}^0\big)\right] && \text{(28c)} \\ &= a^2 + 2a\mathbf{b}^{\mathrm{T}}\big(\boldsymbol{\mu}_{ij} - \mathbf{y}_{ij}^0\big) + \big(\boldsymbol{\mu}_{ij} - \mathbf{y}_{ij}^0\big)^{\mathrm{T}}\big(\mathbf{b}\mathbf{b}^{\mathrm{T}}\big)\big(\boldsymbol{\mu}_{ij} - \mathbf{y}_{ij}^0\big) + \mathrm{Tr}\big(\mathbf{b}\mathbf{b}^{\mathrm{T}}\boldsymbol{\Sigma}_{i,j}\big). && \text{(28d)} \end{aligned}$$

We solve the expectation value g_{S_e} in ④ for an exemplary function $f_{\mathrm{S}_e}\big(\mathbf{y}_{ij}, \mathbf{y}_e\big) = u_{ij} - u_e$. All remaining terms as well as the terms g_{N_e} in ⑥ can be resolved in the same manner. Using $\mathbf{A}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, it holds that

$$\begin{aligned} g_{\mathrm{S}_e}\big(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}, \boldsymbol{\mu}_e, \boldsymbol{\Sigma}_e\big) &= \mathbb{E}_{q(\mathbf{y};\boldsymbol{\theta})}\ f_{\mathrm{S}_e}\big(\mathbf{y}_{ij}, \mathbf{y}_e\big)^2 && \text{(29a)} \\ &= \mathbb{E}_{q(\mathbf{y}_e;\boldsymbol{\theta})}\ \mathbb{E}_{q(\mathbf{y}_{ij};\boldsymbol{\theta})}\left[\big(\mathbf{y}_{ij} - \mathbf{y}_e\big)^{\mathrm{T}}\mathbf{A}_1\big(\mathbf{y}_{ij} - \mathbf{y}_e\big)\right] && \text{(29b)} \\ &= \mathbb{E}_{q(\mathbf{y}_e;\boldsymbol{\theta})}\left[\big(\boldsymbol{\mu}_{ij} - \mathbf{y}_e\big)^{\mathrm{T}}\mathbf{A}_1\big(\boldsymbol{\mu}_{ij} - \mathbf{y}_e\big) + \mathrm{Tr}\big(\mathbf{A}_1\boldsymbol{\Sigma}_{ij}\big)\right] && \text{(29c)} \\ &= \big(\boldsymbol{\mu}_{ij} - \boldsymbol{\mu}_e\big)^{\mathrm{T}}\mathbf{A}_1\big(\boldsymbol{\mu}_{ij} - \boldsymbol{\mu}_e\big) + \mathrm{Tr}\big(\mathbf{A}_1\boldsymbol{\Sigma}_{ij}\big) + \mathrm{Tr}\big(\mathbf{A}_1\boldsymbol{\Sigma}_e\big) && \text{(29d)} \\ &= \Big(\mu_{ij}^{(1)} - \mu_e^{(1)}\Big)^2 + \big(\boldsymbol{\Sigma}_{ij}\big)_{1,1} + \big(\boldsymbol{\Sigma}_e\big)_{1,1} && \text{(29e)} \end{aligned}$$

with $\mu_{ij}^{(1)}$ denoting the first (*i.e.*, horizontal) component of the mean flow vector at pixel (i, j). The term g_{C} in ⑤ can be determined as

$$\begin{aligned} g_{\mathrm{C}}\big(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}, \hat{\boldsymbol{\mu}}_{ij}, \hat{\boldsymbol{\Sigma}}_{ij}\big) &= \mathbb{E}_{q(\mathbf{y},\hat{\mathbf{y}};\boldsymbol{\theta})}\ f_{\mathrm{C}}\big(\mathbf{y}_{ij}, \hat{\mathbf{y}}_{ij}\big)^2 && \text{(30a)} \\ &= \mathbb{E}_{q(\mathbf{y}_{ij};\boldsymbol{\theta})}\ \mathbb{E}_{q(\hat{\mathbf{y}}_{ij};\boldsymbol{\theta})}\left[\big(\mathbf{y}_{ij} - \hat{\mathbf{y}}_{ij}\big)^{\mathrm{T}}\big(\mathbf{y}_{ij} - \hat{\mathbf{y}}_{ij}\big)\right] && \text{(30b)} \\ &= \mathbb{E}_{q(\hat{\mathbf{y}}_{ij};\boldsymbol{\theta})}\left[\big(\boldsymbol{\mu}_{ij} - \hat{\mathbf{y}}_{ij}\big)^{\mathrm{T}}\big(\boldsymbol{\mu}_{ij} - \hat{\mathbf{y}}_{ij}\big) + \mathrm{Tr}\big(\boldsymbol{\Sigma}_{ij}\big)\right] && \text{(30c)} \\ &= \big(\boldsymbol{\mu}_{ij} - \hat{\boldsymbol{\mu}}_{ij}\big)^{\mathrm{T}}\big(\boldsymbol{\mu}_{ij} - \hat{\boldsymbol{\mu}}_{ij}\big) + \mathrm{Tr}\big(\boldsymbol{\Sigma}_{ij}\big) + \mathrm{Tr}\big(\hat{\boldsymbol{\Sigma}}_{ij}\big). && \text{(30d)} \end{aligned}$$

Update equations. To obtain update equations, we compute the derivative of the KL divergence in Eq. (22b), set it to zero, and solve for the desired variable. Please note that update equations for boundary pixels may slightly differ from the ones shown below. From now on, spatial derivatives of I are denoted as I_x and I_y, the temporal derivative is given as I_t. Moreover, $\mathrm{diag}(\cdot)$ represents a diagonal matrix and we define vectors $\mathbf{K}_\gamma = \Big(\sum_l \frac{\mathbf{k}_{\gamma,ij,l}}{\sigma_l^2}\Big)_{ij}$.

As the update of each mean flow estimate $\boldsymbol{\mu}_{ij}$ depends on other entries of $\boldsymbol{\mu}$, it is desirable to jointly solve for all components of the flow field. Therefore, $\boldsymbol{\mu}$ is obtained as the solution of a linear equation system, *c.f.* [50, 37], such that

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} = \Big(\mu_{11}^{(1)}, \ldots, \mu_{nm}^{(1)}, \mu_{11}^{(2)}, \ldots, \mu_{nm}^{(2)}\Big)^{\mathrm{T}}, \quad \mathbf{A} = \mathbf{A}_{\mathrm{D}} + \mathbf{A}_{\mathrm{S}} + \mathbf{A}_{\mathrm{C}}, \quad \mathbf{b} = \mathbf{b}_{\mathrm{D}} + \mathbf{b}_{\mathrm{S}} + \mathbf{b}_{\mathrm{C}}. \tag{31}$$

The components of the linear equation system are determined as

$$\begin{aligned} \mathbf{A}_{\mathrm{D}} &= \lambda_{\mathrm{D}}\begin{pmatrix} \mathrm{diag}\left(\mathbf{K}_{\mathrm{D}}\right)\mathrm{diag}\left(I_x^2\right) & \mathrm{diag}\left(\mathbf{K}_{\mathrm{D}}\right)\mathrm{diag}\left(I_x \cdot I_y\right) \\ \mathrm{diag}\left(\mathbf{K}_{\mathrm{D}}\right)\mathrm{diag}\left(I_x \cdot I_y\right) & \mathrm{diag}\left(\mathbf{K}_{\mathrm{D}}\right)\mathrm{diag}\left(I_y^2\right) \end{pmatrix}, && \text{(32a)} \\ \mathbf{A}_{\mathrm{S}} &= \lambda_{\mathrm{S}}\begin{pmatrix} \mathbf{F}_1^{\mathrm{T}}\,\mathrm{diag}\left(\mathbf{K}_{\mathrm{S}_1}\right)\mathbf{F}_1 + \mathbf{F}_2^{\mathrm{T}}\,\mathrm{diag}\left(\mathbf{K}_{\mathrm{S}_2}\right)\mathbf{F}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_1^{\mathrm{T}}\,\mathrm{diag}\left(\mathbf{K}_{\mathrm{S}_3}\right)\mathbf{F}_1 + \mathbf{F}_2^{\mathrm{T}}\,\mathrm{diag}\left(\mathbf{K}_{\mathrm{S}_4}\right)\mathbf{F}_2 \end{pmatrix}, && \text{(32b)} \\ \mathbf{A}_{\mathrm{C}} &= 2\lambda_{\mathrm{C}}\mathbf{I}, && \text{(32c)} \\ \mathbf{b}_{\mathrm{D}} &= \mathbf{A}_D\mathbf{y}_0 - \lambda_{\mathrm{D}}\begin{pmatrix} \mathrm{diag}\left(\mathbf{K}_{\mathrm{D}}\right)\mathrm{diag}\left(I_x \cdot I_t\right)\mathbf{1} \\ \mathrm{diag}\left(\mathbf{K}_{\mathrm{D}}\right)\mathrm{diag}\left(I_y \cdot I_t\right)\mathbf{1} \end{pmatrix}, \ \mathbf{b}_{\mathrm{S}} = \mathbf{0}, \ \mathbf{b}_{\mathrm{C}} = 2\lambda_{\mathrm{C}}\hat{\boldsymbol{\mu}}. && \text{(32d)} \end{aligned}$$

Here, $\mathbf{F}_1$ and $\mathbf{F}_2$ represent filter matrices corresponding to the derivative filters $\mathbf{H}_1 = \left[1, -1\right]^{\mathrm{T}}$ and $\mathbf{H}_2 = \left[1, -1\right]$, which are used in $f_{\mathrm{S}_e}\left(\mathbf{y}_{ij}, \mathbf{y}_e\right)$, *c.f.* [37]. $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ is a matrix of all zeros.

When updating the auxiliary flow means $\hat{\boldsymbol{\mu}}$, a 5×5 neighborhood has to be considered. Therefore, a joint update of all estimates is computationally expensive and we follow [43] assuming fixed values for neighboring pixels, *i.e.*

$$\hat{\boldsymbol{\mu}}_{ij,t} = \begin{pmatrix} \hat{\mu}^{(1)}_{ij,t} \\ \hat{\mu}^{(2)}_{ij,t} \end{pmatrix}, \quad \hat{\mu}^{(k)}_{ij,t} = \frac{2\lambda_{\mathrm{C}}\, \mu^{(k)}_{ij} + \lambda_{\mathrm{N}} \sum_{e \in N^k(i,j)} \left(\mathbf{K}_{N^k_e}\right)_{ij,t-1} \cdot \hat{\mu}^{(k)}_{e,t-1}}{2\lambda_{\mathrm{C}} + \lambda_{\mathrm{N}} \sum_{e \in N^k(i,j)} \left(\mathbf{K}_{N^k_e}\right)_{ij,t-1}}. \tag{33}$$

Here, $N^k(i,j)$ represents the set of neighbors in terms of the k^{th} optical flow component.

For the flow variances $\mathbf{\Sigma} = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$ we derive a closed-form update dependent only on the latent variables $\mathbf{k}$ with

$$\Sigma_1 = \left(\lambda_{\mathrm{D}} \operatorname{diag}\left(I_x^2\right) \cdot \mathbf{K}_{\mathrm{D}} + \lambda_{\mathrm{S}} \left[\operatorname{abs}\left(\mathbf{F}_1^{\mathrm{T}}\right) \cdot \mathbf{K}_{\mathrm{S}_1} + \operatorname{abs}\left(\mathbf{F}_2^{\mathrm{T}}\right) \cdot \mathbf{K}_{\mathrm{S}_2}\right] + 2\lambda_{\mathrm{C}}\right)^{-1} \tag{34a}$$

and

$$\Sigma_2 = \left(\lambda_{\mathrm{D}} \operatorname{diag}\left(I_y^2\right) \cdot \mathbf{K}_{\mathrm{D}} + \lambda_{\mathrm{S}} \left[\operatorname{abs}\left(\mathbf{F}_1^{\mathrm{T}}\right) \cdot \mathbf{K}_{\mathrm{S}_3} + \operatorname{abs}\left(\mathbf{F}_2^{\mathrm{T}}\right) \cdot \mathbf{K}_{\mathrm{S}_4}\right] + 2\lambda_{\mathrm{C}}\right)^{-1}, \tag{34b}$$

where the absolute value function $\operatorname{abs}(\cdot)$ is applied element-wise.

We assume fixed neighboring values also for the update of the auxiliary flow variances $\hat{\mathbf{\Sigma}}$, and obtain

$$\hat{\mathbf{\Sigma}}_{ij,t} = \begin{pmatrix} \hat{\Sigma}^{(1)}_{ij,t} & 0 \\ 0 & \hat{\Sigma}^{(2)}_{ij,t} \end{pmatrix}, \quad \hat{\Sigma}^{(k)}_{ij,t} = \frac{1}{2\lambda_{\mathrm{C}} + \lambda_{\mathrm{N}} \sum_{e \in N^k(i,j)} \left(\mathbf{K}_{N^k_e}\right)_{ij,t-1}}. \tag{35}$$

To derive an update equation for a latent variable $\mathbf{k}_{\gamma,ij}$, we need to consider a Lagrangian function including the KL divergence in Eq. (22b) as well as the constraint $\sum_l k_{\gamma,ij,l} = 1$. Solving the resulting linear equation system analytically gives us, *e.g.*,

$$k_{\mathrm{D},ij,l} = \left(\frac{\pi_l}{\sigma_l}\right)^{\lambda_{\mathrm{D}}} \exp\left[-\lambda_{\mathrm{D}} \frac{g_{\mathrm{D}}\left(\boldsymbol{\mu}_{ij}, \mathbf{\Sigma}_{ij}; I\right)}{2\sigma_l^2}\right] \cdot Z_{\mathrm{D},ij} \tag{36a}$$

with

$$Z_{\mathrm{D},ij} = \left(\sum_{l=1}^{L} \left(\frac{\pi_l}{\sigma_l}\right)^{\lambda_{\mathrm{D}}} \exp\left[-\lambda_{\mathrm{D}} \frac{g_{\mathrm{D}}\left(\boldsymbol{\mu}_{ij}, \mathbf{\Sigma}_{ij}; I\right)}{2\sigma_l^2}\right]\right)^{-1} \tag{36b}$$

for the latent variables of the data term using the expectation values $g_{\mathrm{D}}\left(\boldsymbol{\mu}_{ij}, \mathbf{\Sigma}_{ij}; I\right)$ as derived in Eq. (28d). Update equations for the remaining latent variables are derived similarly.

B. Bayesian Risk Minimization

We aim to show that the solution of the Bayesian risk minimization in Eq. (8) is given as $\mathbf{y}^\star_{ij} = \boldsymbol{\mu}_{ij}$ when replacing the posterior p with its approximating distribution q and using the Average End-Point Error (AEPE) as a loss function.

Recall that the AEPE is defined as $l\left(\mathbf{y}, \tilde{\mathbf{y}}\right) = \sum_{i,j} \ell\left(\mathbf{y}_{ij}, \tilde{\mathbf{y}}_{ij}\right) = \sum_{i,j} \|\mathbf{y}_{ij} - \tilde{\mathbf{y}}_{ij}\|_2$ with

$$\begin{aligned} \nabla_2\, \ell\left(\mathbf{a} - \mathbf{x}, \tilde{\mathbf{x}}\right) &= \left(\mathbf{a} - \mathbf{x} - \tilde{\mathbf{x}}\right) / \|\mathbf{a} - \mathbf{x} - \tilde{\mathbf{x}}\|_2 && (37a) \\ &= -\left(\mathbf{x} - \left(\mathbf{a} - \tilde{\mathbf{x}}\right)\right) / \|\mathbf{x} - \left(\mathbf{a} - \tilde{\mathbf{x}}\right)\|_2 && (37b) \\ &= -\nabla_2\, \ell\left(\mathbf{x}, \mathbf{a} - \tilde{\mathbf{x}}\right) && (37c) \end{aligned}$$

for arbitrary $\mathbf{a} \in \mathbb{R}^2$. W.l.o.g. we minimize the expected risk of $l(\mathbf{y}, \tilde{\mathbf{y}})$ and therefore set $f(\tilde{\mathbf{y}}) = \mathbb{E}_{q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}; \boldsymbol{\theta})}\left[l(\mathbf{y}, \tilde{\mathbf{y}})\right]$. Note that we omit the variational parameters $\boldsymbol{\theta}$ in the following for brevity. Using the properties of q, we obtain

$$\begin{aligned} f(\tilde{\mathbf{y}}) &= \int_{\mathcal{Y}} \int_{\hat{\mathcal{Y}}} \sum_{\mathcal{H}} q(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{h}) \cdot l\left(\mathbf{y}, \tilde{\mathbf{y}}\right) d\mathbf{y}\; d\hat{\mathbf{y}} && (38a) \\ &\overset{\text{q fac.}}{=} \int_{\mathcal{Y}} q(\mathbf{y}) \cdot l\left(\mathbf{y}, \tilde{\mathbf{y}}\right) d\mathbf{y} && (38b) \\ &= \sum_{i,j} \underbrace{\int_{\mathbb{R}^2} q(\mathbf{y}_{ij}) \cdot \ell\left(\mathbf{y}_{ij}, \tilde{\mathbf{y}}_{ij}\right) d\mathbf{y}_{ij}}_{=: f_{ij}(\tilde{\mathbf{y}}_{ij})}. && (38c) \end{aligned}$$

For fixed $\mathbf{y}_{ij} \in \mathbb{R}^2$, the function $q(\mathbf{y}_{ij}) \cdot \ell(\mathbf{y}_{ij}, \tilde{\mathbf{y}}_{ij})$ is convex in $\tilde{\mathbf{y}}_{ij}$. Therefore, the objective $f(\tilde{\mathbf{y}})$ is convex in $\tilde{\mathbf{y}}$ and the Bayesian risk minimization has a unique solution given by

$$\mathbf{y}_{ij} = \arg\min_{\tilde{\mathbf{y}}_{ij}} f_{ij}(\tilde{\mathbf{y}}_{ij}). \tag{39}$$

It only remains to be shown that $\nabla_2 f_{ij}(\tilde{\mathbf{y}}_{ij}) = 0$ holds for $\tilde{\mathbf{y}}_{ij} = \boldsymbol{\mu}_{ij}$. Setting $\tilde{\mathbf{y}}_{ij} = \boldsymbol{\mu}_{ij}$ we obtain

$$\begin{aligned}
\int_{-\infty}^{\boldsymbol{\mu}_{ij}} q(\boldsymbol{\tau}) \cdot \nabla_2\, \ell\left(\boldsymbol{\tau}, \boldsymbol{\mu}_{ij}\right) d\boldsymbol{\tau} &\overset{(\mathbf{z}_1 = \boldsymbol{\tau} - \boldsymbol{\mu}_{ij})}{=} \int_{-\infty}^{\mathbf{0}} q(\boldsymbol{\mu}_{ij} + \mathbf{z}_1) \cdot \nabla_2\, \ell\left(\boldsymbol{\mu}_{ij} + \mathbf{z}_1, \boldsymbol{\mu}_{ij}\right) d\mathbf{z}_1 && \text{(40a)} \\
&\overset{q \text{ sym.}}{=} \int_{-\infty}^{\mathbf{0}} q(\boldsymbol{\mu}_{ij} - \mathbf{z}_1) \cdot \nabla_2\, \ell\left(\boldsymbol{\mu}_{ij} + \mathbf{z}_1, \boldsymbol{\mu}_{ij}\right) d\mathbf{z}_1 && \text{(40b)} \\
&\overset{(\mathbf{z}_2 = \boldsymbol{\mu}_{ij} - \mathbf{z}_1)}{=} \int_{\boldsymbol{\mu}_{ij}}^{\infty} q(\mathbf{z}_2) \cdot \nabla_2\, \ell\left(2\boldsymbol{\mu}_{ij} - \mathbf{z}_2, \boldsymbol{\mu}_{ij}\right) d\mathbf{z}_2 && \text{(40c)} \\
&\overset{(37a)-(37c)}{=} -\int_{\boldsymbol{\mu}_{ij}}^{\infty} q(\mathbf{z}_2) \cdot \nabla_2\, \ell\left(\mathbf{z}_2, \boldsymbol{\mu}_{ij}\right) d\mathbf{z}_2 && \text{(40d)}
\end{aligned}$$

and finally

$$\begin{aligned}
\nabla_2\, f_{ij}\left(\boldsymbol{\mu}_{ij}\right) &= \int_{\mathbb{R}^2} q(\boldsymbol{\tau}) \cdot \nabla_2\, \ell\left(\boldsymbol{\tau}, \boldsymbol{\mu}_{ij}\right) d\boldsymbol{\tau} && \text{(41a)} \\
&= \int_{-\infty}^{\boldsymbol{\mu}_{ij}} q(\boldsymbol{\tau}) \cdot \nabla_2\, \ell\left(\boldsymbol{\tau}, \boldsymbol{\mu}_{ij}\right) d\boldsymbol{\tau} + \int_{\boldsymbol{\mu}_{ij}}^{\infty} q(\boldsymbol{\tau}) \cdot \nabla_2\, \ell\left(\boldsymbol{\tau}, \boldsymbol{\mu}_{ij}\right) d\boldsymbol{\tau} \\
&\overset{(40d)}{=} -\int_{\boldsymbol{\mu}_{ij}}^{\infty} q(\boldsymbol{\tau}) \cdot \nabla_2\, \ell\left(\boldsymbol{\tau}, \boldsymbol{\mu}_{ij}\right) d\boldsymbol{\tau} + \int_{\boldsymbol{\mu}_{ij}}^{\infty} q(\boldsymbol{\tau}) \cdot \nabla_2\, \ell\left(\boldsymbol{\tau}, \boldsymbol{\mu}_{ij}\right) d\boldsymbol{\tau} && \text{(41b)} \\
&= 0. && \text{(41c)}
\end{aligned}$$

C. Implementation Details

In this section, we present our design choices following the best-practices of energy-based optical flow techniques, and give an analysis evaluating the influence of the specifics. Moreover, we give details of our post-processing approach using the fast bilateral solver [4].

C.1. ProbClassicA

In our ProbClassicA algorithm, we perform three steps of graduated non-convexity and apply coarse-to-fine estimation with 10 warping steps per layer. As in [43], we restrict the flow update to an absolute value of 1 and pre-process the images using a structure-texture decomposition. Spline-based cubic interpolation as well as an averaging of image gradients $\nabla_2 I_1$ and $\nabla_2 I_2$ are applied. During the inference, the variable sets $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{k}\}$ and $\{\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \hat{\mathbf{k}}\}$ are updated in an alternating way. As an inner update step, we apply five iterations of the block-coordinate descent scheme on $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\mathbf{k}$. For the set $\{\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \hat{\mathbf{k}}\}$, a number of three inner updates performs better.

C.2. ProbFlowFields

For ProbFlowFields, we follow [36] and pre-smooth images using a Gaussian kernel of size 9×9 with $\sigma = 1.1$. For warping, we apply bilinear interpolation and averaged image derivatives. Moreover, we perform five warping steps, each with five iterations of our block-coordinate descent scheme. We follow Revaud *et al.* and compute optical flow updates with 30 iterations of successive over relaxation, which performs noticeably faster than the solver used in [43].

C.3. Evaluation of design choices

Table 6 summarizes results of AEPE, AUC, and CC on the Middlebury and Sintel benchmarks using varying setups of ProbClassicA and ProbFlowFields. In a first step, we evaluate a setting for ProbClassicA in which parameters λ_{D}, λ_{S}, and λ_{N} are determined by having the Bayesian optimization [42] consider only the AEPE or only the AUC instead of the F_1-score

ProbClassicA Middlebury	AEPE	rel. chg.	AUC	rel. chg.	CC	rel. chg.
Baseline	0.296	–	0.466	–	0.374	–
Bayesian optim. w.r.t. AEPE only	0.290	**-0.02**	0.471	0.01	0.351	**0.06**
Bayesian optim. w.r.t. AUC only	0.312	**0.05**	0.436	**-0.06**	0.451	**-0.21**
$E_{\mathrm{N}} = E_{\mathrm{C}} = 0$	0.411	**0.39**	0.889	**0.91**	0.125	**0.67**
No structure-texture decomposition	0.290	**-0.02**	0.445	**-0.05**	0.361	**0.03**
ProbFlowFields Sintel validation	**AEPE**	**rel. chg.**	**AUC**	**rel. chg.**	**CC**	**rel. chg.**
Baseline	3.127	–	0.398	–	0.563	–
Bayesian optim. w.r.t. AEPE only	3.128	<0.01	0.475	**0.19**	0.407	**0.28**
Bayesian optim. w.r.t. AUC only	3.219	**0.03**	0.381	**-0.04**	0.644	**-0.14**
Spatially constant λ_{S}	3.127	0.00	0.400	<0.01	0.562	<0.01
$\theta_{ij}^{r} = 1$	3.125	>-0.01	0.396	>-0.01	0.548	**0.03**
No gradient averaging	3.135	<0.01	0.398	0.00	0.557	0.01
No Gaussian smoothing	3.135	<0.01	0.441	**0.11**	0.497	**0.12**
10 warping steps	3.123	>-0.01	0.421	**0.06**	0.538	**0.04**

Table 6. Analysis of several design choices for ProbClassicA on Middlebury and ProbFlowFields on the Sintel validation set. Bold entries denote strong deviations from the baseline.

proposed in Eq. (20). In both cases, we observe that the performance w.r.t. the evaluation metric that is not considered during the Bayesian optimization drops significantly. This highlights the importance of the F_1-score to balance the accuracy of flow and uncertainty estimates. Moreover, we show that the AEPE as well as the performance of the uncertainty measure is clearly inferior if no additional nonlocal term is applied ($E_{\mathrm{N}} = E_{\mathrm{C}} = 0$). When using ProbClassicA without structure-texture decomposition as pre-processing, we surprisingly obtain improved results for the AEPE (2%) as well as the AUC (5%). This is in contrast to energy minimization, where this pre-processing helps [43]. For fairness of comparison to the underlying energy minimization approach, we continue to use a structure-texture decomposition.

Considering ProbFlowFields, we observe the same behavior as for ProbClassicA when Bayesian optimization is carried out only with respect to one of the evaluation metrics. Note that the parameter setting obtained by a Bayesian optimization w.r.t. to the AEPE performs better than the baseline on the training set even though no improvement of the AEPE is visible on the validation set. The usage of a spatially constant trade-off parameter λ_{S}, turning off the normalization of the spatial derivatives ($\theta_{ij}^{r} = 1$, *c.f.* Eqs. (17) and (18)), and not averaging the image gradients, respectively, only lead to minor changes. When no Gaussian smoothing is applied for image pre-processing, a clear effect on the AUC as well as the CC can be observed whereas the AEPE is only slightly changed. Finally, the application of 10 warping steps only results in small improvements of the AEPE and even decreases the performance of the uncertainty measure. This justifies the usage of a reduced number of 5 steps to save computational time.

C.4. Post-processing using the fast bilateral solver

As described in Sec. 7.3, we apply the fast bilateral solver [4] on top of ProbFlowFields in order to illustrate the benefits of uncertainty predictions for a further improvement of the flow estimates. In doing so, we normalize the estimated uncertainties with a sigmoid function and invert the values to obtain the confidences required by the fast bilateral solver. A Bayesian optimization [42] is performed on our Sintel training set to obtain appropriate sigmoid parameters as well as a suitable trade-off parameter for the fast bilateral solver. See Fig. 5 for a screenshot of the private Sintel benchmark table showing results after post-processing (ProbFlowFields + BS). For the reported baseline, we process the estimates of ProbFlowFields assuming a uniform confidence of 0.5.

D. Additional Uncertainty Measures

In the following, we evaluate several additional uncertainty measures on the Middlebury as well as the Sintel benchmark. Haußecker and Spies [20] introduce three confidence measures based on the spatio-temporal structure tensor

$$\mathbf{S} = G(\tilde{\sigma}) * \left[(\nabla_3 I)(\nabla_3 I)^{\mathrm{T}}\right] \qquad \text{with} \qquad \nabla_3 I = (I_x, I_y, I_t)^{\mathrm{T}}, \tag{42}$$

where I_x and I_y denote the spatial image derivatives computed with central differences and I_t is the temporal difference between I_1 and I_2. Following [30], we smooth the derivatives with a Gaussian filter $G(\tilde{\sigma})$ of size 7×7 and a standard

	EPE all	EPE matched	EPE unmatched	d0-10	d10-60	d60-140	s0-10	s10-40	s40+	
GroundTruth [1]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	Visualize Results
DCFlow [2]	5.119	2.283	28.228	4.665	2.108	1.440	1.052	3.434	29.351	Visualize Results
FlowFieldsCNN [3]	5.363	2.303	30.313	4.718	2.020	1.399	1.032	3.065	32.422	Visualize Results
MR-Flow [4]	5.376	2.818	26.235	5.109	2.395	1.755	0.908	3.443	32.221	Visualize Results
FTFlow [5]	5.390	2.268	30.841	4.513	1.964	1.366	1.046	3.322	31.936	Visualize Results
S2F-IF [6]	5.417	2.549	28.795	4.745	2.198	1.712	1.157	3.468	31.262	Visualize Results
InterpoNet_ff [7]	5.535	2.372	31.296	4.720	2.018	1.532	1.064	3.496	32.633	Visualize Results
RegionalFF [8]	5.562	2.595	29.741	4.921	2.393	1.639	1.122	3.477	32.625	Visualize Results
PGM-C [9]	5.591	2.672	29.389	4.975	2.340	1.791	1.057	3.421	33.339	Visualize Results
RicFlow [10]	5.620	2.765	28.907	5.146	2.366	1.679	1.088	3.364	33.573	Visualize Results
InterpoNet_cpm [11]	5.627	2.594	30.344	4.975	2.213	1.640	1.042	3.575	33.321	Visualize Results
ProbFlowFields+BS [12]	5.628	2.543	30.773	4.680	2.169	1.683	1.086	3.538	33.210	Visualize Results
CPM_AUG [13]	5.645	2.737	29.362	4.707	2.150	1.918	1.087	3.306	33.925	Visualize Results
ProbFlowFields [14]	5.696	2.545	31.371	4.696	2.150	1.686	1.146	3.658	33.188	Visualize Results

Figure 5. Screenshot of private Sintel table (final) showing results for ProbFlowFields and ProbFlowFields + BS (status as of July 2017).

Uncertainty measure	AUC	rel. chg.	CC	rel. chg.
Ct [20]	1.058	1.27	-0.106	1.28
Cs [20]	1.014	1.18	-0.057	1.15
Cc [20]	0.967	1.08	-0.022	1.06
Ev3 [27]	0.989	1.12	0.058	0.84
Noise	0.512	0.10	0.286	0.24
ProbClassicA *(ours)*	**0.466**	**0.00**	**0.374**	**0.00**
Oracle	*0.255*	–	*1.000*	–

Table 7. Area under curve (AUC), Spearman's rank correlation coefficient (CC), and relative change (rel. chg.) in comparison to the our uncertainty measure on the Middlebury dataset.

Uncertainty measure	AUC	rel. chg.	CC	rel. chg.
Ct [20]	1.130	1.84	-0.128	1.23
Cs [20]	1.154	1.90	-0.149	1.26
Cc [20]	0.915	1.30	0.129	0.77
Ev3 [27]	1.024	1.57	-0.030	1.05
Noise	0.512	0.29	0.382	0.32
ProbFlowFields *(ours)*	**0.398**	**0.00**	**0.563**	**0.00**
Oracle	*0.182*	–	*1.000*	–

Table 8. Area under curve (AUC), Spearman's rank correlation coefficient (CC), and relative change (rel. chg.) in comparison to our uncertainty measure on a Sintel benchmark validation set.

deviation $\tilde{\sigma} = 2$. In [20], eigenvalues λ_1, λ_2, and λ_3 of $\mathbf{S}$ are computed such that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Uncertainty measures are then obtained as

$$\Psi_{\mathrm{Ct}} = -\left(\frac{\lambda_1 - \lambda_3}{\lambda_1 + \lambda_3}\right)^2, \quad \Psi_{\mathrm{Cs}} = -\left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2, \quad \text{and} \quad \Psi_{\mathrm{Cc}} = \Psi_{\mathrm{Ct}} - \Psi_{\mathrm{Cs}}. \tag{43}$$

Moreover, we evaluate a baseline uncertainty measure as used in [27] defined as $\Psi_{\mathrm{Ev3}} = -\lambda_3$.

Finally, we compare to a sampling-based measure similar to the idea of Kybic and Nieuwenhuis [27]. That is, we estimate the uncertainty as the variance of the optical flow estimates resulting from small, random perturbations of the input data. Specifically, we apply zero-mean Gaussian noise on the input images and determine appropriate values for the variance of the noise on the training set. The uncertainty measure is then obtained as $\Psi_{\mathrm{Noise}} = \sqrt{\sigma_u^2 + \sigma_v^2}$ with σ_u and σ_v denoting the standard derivation of the horizontal and vertical flow estimates per pixel.

As can be seen in Tables 7 and 8, all measures based on the structure tensor perform considerably worse than our proposed uncertainty measure. Ψ_{Cc} and Ψ_{Ev3} lead to more meaningful uncertainties than the two remaining approaches on both datasets, but perform similar to the simple gradient-based measure [3]. The noise uncertainty – especially on the Middlebury dataset – performs comparably to Ψ_{Energy} and Ψ_{Learned}. However, our ProbFlow approach clearly leads to superior results.

E. ProbFlowFields on Middlebury

For completeness, we report the results of ProbFlowFields on Middlebury. To reproduce the Middlebury results shown in [1] we applied the default settings of the EpicFlow interpolation. Moreover, we use GSM potentials trained on the Sintel

Method	training AEPE	training rel. chg.	test AEPE	test rel. chg.
Initialization	0.307	0.38	–	–
FlowFields [1]	0.240	0.08	0.331†	0.10
FieldsFields*	0.230	0.04	–	–
ProbFlowFields *(ours)*	**0.222**	**0.00**	**0.301**	**0.00**

Table 9. Average end-point error (AEPE) and relative change (rel. chg.) in comparison to the ProbFlowFields method on the Middlebury benchmark. †Please note that we did not re-evaluate FlowFields, but show the publicly available results.

Uncertainty measure	AUC	rel. chg.	CC	rel. chg.
Gradient [3]	1.244	1.72	-0.077	1.21
Laplace	0.539	0.18	0.297	0.20
Energy [8]	0.563	0.23	0.253	0.32
Learned [30]	0.473	0.04	**0.374**	**>-0.01**
ProbFlowFields *(ours)*	**0.457**	**0.00**	0.371	0.00
Oracle	*0.247*	–	*1.000*	–

Table 10. Area under curve (AUC), Spearman's rank correlation coefficient (CC), and relative change (rel. chg.) in comparison to the energy uncertainty measure on the Middlebury training set.

Average endpoint error	avg. rank	Army (Hidden texture) GT im0 im1 all	disc	untext	Mequon (Hidden texture) GT im0 im1 all	disc	untext	Schefflera (Hidden texture) GT im0 im1 all	disc	untext	Wooden (Hidden texture) GT im0 im1 all	disc	untext	Grove (Synthetic) GT im0 im1 all	disc	untext	Urban (Synthetic) GT im0 im1 all	disc	untext	Yosemite (Synthetic) GT im0 im1 all	disc	untext	Teddy (Stereo) GT im0 im1 all	disc	untext
NNF-Local [87]	3.4	0.07 1	0.20 2	0.05 1	0.15 2	0.51 4	0.12 5	0.18 2	0.37 2	0.14 2	0.10 2	0.49 5	0.06 2	0.41 1	0.61 1	0.21 2	0.23 2	0.66 3	0.19 2	0.10 6	0.12 12	0.17 13	0.34 2	0.80 6	0.23 2
PMMST [114]	9.3	0.09 32	0.21 5	0.07 16	0.18 11	0.51 4	0.16 28	0.21 9	0.42 8	0.17 16	0.10 2	0.33 1	0.08 13	0.51 5	0.74 4	0.28 7	0.24 3	0.65 2	0.20 5	0.11 18	0.12 12	0.17 13	0.37 4	0.74 2	0.35 4
OFLAF [77]	9.8	0.08 9	0.21 5	0.06 6	0.16 6	0.53 6	0.12 5	0.19 3	0.37 2	0.14 2	0.14 9	0.77 29	0.07 5	0.51 5	0.78 7	0.25 4	0.31 11	0.76 4	0.25 16	0.11 18	0.12 12	0.21 40	0.42 10	0.78 4	0.63 18
MDP-Flow2 [68]	10.6	0.08 9	0.21 5	0.07 16	0.15 2	0.48 1	0.11 1	0.20 5	0.40 5	0.14 2	0.15 22	0.80 37	0.08 13	0.63 19	0.93 19	0.43 20	0.26 5	0.76 4	0.23 9	0.11 18	0.12 12	0.17 13	0.38 6	0.79 5	0.44 6
NN-field [71]	11.8	0.08 9	0.22 17	0.05 1	0.17 8	0.55 10	0.13 11	0.19 3	0.39 4	0.15 7	0.09 1	0.48 4	0.05 1	0.41 1	0.61 1	0.20 1	0.52 60	0.64 1	0.26 19	0.13 42	0.13 36	0.20 34	0.35 3	0.83 8	0.21 1
ComponentFusion [96]	13.9	0.07 1	0.21 5	0.05 1	0.16 6	0.55 10	0.12 5	0.20 5	0.44 9	0.15 7	0.11 4	0.65 9	0.06 2	0.71 35	1.07 40	0.53 37	0.32 15	1.06 27	0.28 22	0.11 18	0.13 36	0.15 8	0.41 9	0.88 13	0.54 9
TC/T-Flow [76]	19.5	0.07 1	0.21 5	0.05 1	0.19 18	0.68 34	0.12 5	0.28 26	0.66 31	0.14 2	0.14 9	0.86 47	0.07 5	0.67 29	0.98 29	0.49 31	0.22 1	0.82 9	0.19 2	0.11 18	0.11 2	0.30 85	0.50 27	1.02 30	0.64 20
WLIF-Flow [93]	19.8	0.08 9	0.21 5	0.06 6	0.18 11	0.55 10	0.15 23	0.25 18	0.56 20	0.17 16	0.14 9	0.68 10	0.08 13	0.61 16	0.91 17	0.41 18	0.43 35	0.96 15	0.29 28	0.13 42	0.12 12	0.21 40	0.51 32	1.03 33	0.72 38
NNF-EAC [103]	21.3	0.09 32	0.22 17	0.07 16	0.17 8	0.53 6	0.13 11	0.23 11	0.49 12	0.15 7	0.16 36	0.80 37	0.09 28	0.60 13	0.89 13	0.40 16	0.38 25	0.78 6	0.28 22	0.12 31	0.12 12	0.18 26	0.57 45	1.24 49	0.69 32
Layers++ [37]	21.9	0.08 9	0.21 5	0.07 16	0.19 18	0.56 13	0.17 35	0.20 5	0.40 5	0.18 27	0.13 8	0.58 7	0.07 5	0.48 3	0.70 3	0.33 9	0.47 47	1.01 19	0.33 47	0.15 65	0.14 58	0.24 53	0.46 17	0.88 13	0.72 38
LME [70]	22.9	0.08 9	0.22 17	0.06 6	0.15 2	0.49 2	0.11 1	0.30 35	0.64 26	0.31 89	0.15 22	0.78 33	0.09 28	0.66 25	0.96 24	0.53 37	0.33 16	1.18 44	0.28 22	0.12 31	0.12 12	0.18 26	0.44 12	0.91 16	0.61 14
IROF++ [58]	23.0	0.08 9	0.23 24	0.07 16	0.21 32	0.68 34	0.17 35	0.28 26	0.63 24	0.19 39	0.15 22	0.73 22	0.09 28	0.60 13	0.89 13	0.42 19	0.43 35	1.08 30	0.31 37	0.10 6	0.12 12	0.12 4	0.47 19	0.98 23	0.68 31
nLayers [57]	23.8	0.07 1	0.19 1	0.06 6	0.22 40	0.59 16	0.19 57	0.25 18	0.54 16	0.20 48	0.15 22	0.84 43	0.08 13	0.53 7	0.78 7	0.34 11	0.44 39	0.84 10	0.30 33	0.13 42	0.13 36	0.20 34	0.47 19	0.97 22	0.67 29
HAST [109]	25.1	0.07 1	0.20 2	0.05 1	0.18 11	0.54 8	0.13 11	0.17 1	0.32 1	0.12 1	0.15 22	0.90 58	0.06 2	0.49 4	0.74 4	0.22 3	0.58 70	1.09 31	0.44 70	0.19 96	0.17 87	0.47 114	0.32 1	0.64 1	0.33 3
PH-Flow [101]	25.8	0.08 9	0.24 31	0.07 16	0.21 32	0.68 34	0.17 35	0.23 11	0.49 12	0.19 39	0.16 36	0.83 41	0.09 28	0.56 9	0.83 9	0.38 13	0.30 9	0.81 7	0.24 14	0.15 65	0.13 36	0.30 85	0.43 11	0.85 9	0.66 27
FC-2Layers-FF [74]	25.8	0.08 9	0.21 5	0.07 16	0.21 32	0.70 42	0.17 35	0.20 5	0.40 5	0.18 27	0.15 22	0.76 28	0.08 13	0.53 7	0.77 6	0.37 12	0.49 53	1.02 20	0.33 47	0.16 76	0.13 36	0.29 80	0.44 12	0.87 12	0.64 20
Correlation Flow [75]	26.5	0.09 32	0.23 24	0.07 16	0.17 8	0.58 15	0.11 1	0.43 65	0.99 67	0.15 7	0.11 4	0.47 3	0.08 13	0.75 41	1.08 41	0.56 42	0.41 31	0.92 13	0.30 33	0.14 52	0.13 36	0.27 69	0.40 8	0.85 9	0.42 5
AGIF+OF [85]	28.0	0.08 9	0.22 17	0.07 16	0.23 55	0.73 47	0.18 47	0.28 26	0.66 31	0.18 27	0.14 9	0.70 13	0.08 13	0.57 10	0.85 10	0.38 13	0.47 47	0.97 16	0.31 37	0.13 42	0.13 36	0.22 46	0.51 32	0.99 26	0.74 47
RNLOD-Flow [121]	28.0	0.07 1	0.20 2	0.06 6	0.19 18	0.68 34	0.13 11	0.33 49	0.79 50	0.17 16	0.14 9	0.73 22	0.07 5	0.69 33	1.03 33	0.48 27	0.37 24	0.99 17	0.29 28	0.16 76	0.16 79	0.29 80	0.45 14	0.88 13	0.65 25
ProbFlowFields [128]	28.8	0.10 47	0.31 73	0.08 44	0.19 18	0.63 23	0.17 35	0.27 22	0.63 24	0.22 59	0.11 4	0.49 5	0.07 5	0.82 51	1.22 56	0.59 45	0.25 4	1.05 26	0.21 6	0.09 4	0.12 12	0.17 13	0.58 47	1.33 51	0.62 16
⋮											⋮													⋮	
DeepFlow [86]	65.9	0.12 79	0.31 73	0.11 91	0.28 81	0.82 72	0.22 82	0.44 71	1.00 68	0.33 90	0.26 86	1.34 96	0.15 81	0.81 49	1.21 53	0.58 44	0.38 25	1.55 78	0.25 16	0.11 18	0.11 2	0.24 53	0.93 91	1.82 93	1.12 89
ProbClassicA [127]	66.5	0.10 47	0.28 54	0.08 44	0.22 40	0.78 61	0.16 28	0.57 88	1.17 84	0.22 59	0.21 72	1.24 86	0.11 56	0.86 58	1.30 72	0.63 52	0.55 65	1.74 93	0.37 66	0.16 76	0.14 58	0.34 99	0.81 82	1.74 85	0.93 71
TriangleFlow [30]	66.7	0.11 59	0.29 62	0.09 65	0.26 73	0.95 88	0.17 35	0.47 74	1.07 72	0.18 27	0.16 36	0.87 51	0.09 28	1.07 97	1.47 104	1.10 98	0.87 86	1.39 65	0.57 87	0.15 65	0.19 109	0.23 52	0.63 55	1.33 51	0.84 61

Figure 6. Screenshot of private Middlebury table showing results for ProbFlowFields and ProbClassicA (status as of July 2017).

dataset for our ProbFlowFields approach. The results evaluating the AEPE on the Middlebury benchmark can be found in Table 9. We outperform the original FlowFields approach on training and test and obtain improved results in comparison to FlowFields*. Please note that the Middlebury benchmark policy allows no more than one entry per method in the public table. Therefore, we decided to show the results of ProbFlowFields on the Middlebury website whereas the results of ProbClassicA from Table 1 of the main paper are only visible in a private table, see Fig. 6 for a screenshot.

Table 10 shows an evaluation of different uncertainty measures. In contrast to our remaining experiments, the Laplace and learned uncertainty measures both outperform the energy-based approach. Our uncertainty measure is slightly outperformed by Ψ_{Learned} w.r.t. the CC metric. However, ProbFlowFields shows clearly superior results considering the AUC.

References

[50] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. *ECCV*, 2004. 3