Parametric Object Motion from Blur
– Supplementary Material –

Jochen Gast Anita Sellent Stefan Roth
Department of Computer Science, TU Darmstadt

In this supplementary material we show how to derive the free energy as well as the necessary update equations. Moreover, we present additional experimental results.

Notation. In the following we will make use of an overloaded notation for both indexing vectors and concatenating scalars (into vectors). That is, whenever a vector $\mathbf{f}$ is given, we retrieve its i^{th} element via $(\mathbf{f})_i$. On the other hand, we concatenate scalar elements f_i into the vector $(f_i)_i$.

A. Free Energy

We begin by deriving the free energy as stated in Eq. (14) of the main paper. For conciseness, we express some of the frequently appearing integrals directly as expectations, *i.e.*

$$\int q(\mathbf{x})\phi(\mathbf{x})\,\mathrm{d}\mathbf{x} \equiv \langle \phi(\mathbf{x}) \rangle_{q(\mathbf{x})}. \tag{18}$$

Given the independence assumptions of $q(\mathbf{x}, \mathbf{h}, \mathbf{l})$ in Eq. (12) we can thus rewrite Eq. (13) as

$$\begin{aligned} F(q, \mathbf{a}) = &- \langle \log p(\mathbf{y} \,|\, \mathbf{x}, \mathbf{h}, \mathbf{a}) \rangle_{q(\mathbf{x},\mathbf{h})} \\ &- \langle \log p(\mathbf{x}, \mathbf{l}) \rangle_{q(\mathbf{x},\mathbf{l})} - \langle \log p(\mathbf{h}) \rangle_{q(\mathbf{h})} \\ &+ \langle \log q(\mathbf{x}) \rangle_{q(\mathbf{x})} + \langle \log q(\mathbf{h}) \rangle_{q(\mathbf{h})} \\ &+ \sum_{i,\gamma} \langle \log q(\mathbf{l}_{i,\gamma}) \rangle_{q(\mathbf{l}_{i,\gamma})}. \end{aligned} \tag{19}$$

As derived in [8, 18, 25], the entropy terms of the approximating distribution simplify to

$$\langle \log q(\mathbf{x}) \rangle_{q(\mathbf{x})} = -\frac{1}{2} \sum_i \log(\boldsymbol{\sigma}_{\mathbf{x}})_i + \text{const}, \tag{20}$$

$$\langle \log q(\mathbf{h}) \rangle_{q(\mathbf{h})} = \sum_i r_i \log r_i + (1 - r_i) \log(1 - r_i), \tag{21}$$

$$\langle \log q(\mathbf{l}_{i,\gamma}) \rangle_{q(\mathbf{l}_{i,\gamma})} = \sum_j v_{i,\gamma,j} \log v_{i,\gamma,j}. \tag{22}$$

The term involving the sparse image prior is similarly derived in [18] and simplifies to

$$\begin{aligned} -\langle \log p(\mathbf{x}, \mathbf{l}) \rangle_{q(\mathbf{x},\mathbf{l})} = &\left\langle \sum_{i,\gamma,j} v_{i,\gamma,j} \tfrac{\|f_{i,\gamma}(\mathbf{x})\|^2}{2\sigma_j^2} \right\rangle_{q(\mathbf{x})} \\ &+ \sum_{i,\gamma,j} v_{i,\gamma,j} (\log \sigma_j - \log \pi_j) \\ &+ \text{const}. \end{aligned} \tag{23}$$

To simplify further, we rely on each derivative filter $f_{i,\gamma}$ in Eq. (23) corresponding to a linear operator $\mathbf{D}_\gamma$, *i.e.*

$$f_{i,\gamma}(\mathbf{x}) = (\mathbf{D}_\gamma \mathbf{x})_i. \tag{24}$$

Inserting Eq. (24) into Eq. (23) and expanding the expectation, we can explicitly express it in terms of the moments $(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\sigma}_{\mathbf{x}})$ of $q(\mathbf{x})$:

$$\begin{aligned} -\langle \log p(\mathbf{x}, \mathbf{l}) \rangle_{q(\mathbf{x},\mathbf{l})} = &\\ &\sum_{i,\gamma,j} \frac{v_{i,\gamma,j}}{2\sigma_j^2} \left(\mathbf{D}_\gamma \boldsymbol{\mu}_{\mathbf{x}} \circ \mathbf{D}_\gamma \boldsymbol{\mu}_{\mathbf{x}} + (\mathbf{D}_\gamma \circ \mathbf{D}_\gamma) \boldsymbol{\sigma}_{\mathbf{x}} \right)_i \\ &+ \sum_{i,\gamma,j} v_{i,\gamma,j} (\log \sigma_j - \log \pi_j) + \text{const}. \end{aligned} \tag{25}$$

where $\circ$ denotes the pointwise Hadamard product.

Using the prior on the segmentation as defined in Section 4 of the main paper we obtain

$$\begin{aligned} -\langle \log p(\mathbf{h}) \rangle_{q(\mathbf{h})} = \text{const} - &\left\langle \sum_i -\lambda_0 h_i \right. \\ &\left. + \sum_{(i,j) \in N} -\lambda \left[h_i \neq h_j \right] \right\rangle_{q(\mathbf{h})} \\ = \lambda_0 \sum_i r_i + \lambda &\sum_{(i,j) \in N} r_i + r_j - 2 r_i r_j \\ &+ \text{const}. \end{aligned} \tag{26}$$

Note that Eq. (26) extends [25] by including the additional bias term $\left(\lambda_0 \sum_i r_i\right)$ favoring a background segmentation. The last term to be derived involves the expected

log-likelihood

$$-\left\langle \log p(\mathbf{y}\,|\,\mathbf{x},\mathbf{h},\mathbf{a})\right\rangle_{q(\mathbf{x},\mathbf{h})} = \text{const} + \int q(\mathbf{x})q(\mathbf{h})\,\frac{\|\mathbf{h}\circ(\mathbf{K}^{\mathbf{a}}\mathbf{x})+(\mathbf{1}-\mathbf{h})\circ\mathbf{x}-\mathbf{y}\|^2}{2\sigma_n^2}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{h}, \quad (27)$$

induced by the Gaussian noise assumption of Eq. (9). In order to expand the expectation in Eq. (27), we face the challenge that the latent image $\mathbf{x}$ has a larger domain than the blurry input image $\mathbf{y}$. For this reason we introduce a crop operator $\mathbf{I}_\mathbf{y}$ that maps pixel positions of $\mathbf{x}$ to positions in $\mathbf{y}$. Inserting the crop operator and utilizing standard formulas to compute expected values of quadratic norms (see, *e.g.*, [41]) yields the expected log-likelihood

$$\begin{aligned}-\left\langle \log p(\mathbf{y}\,|\,\mathbf{x},\mathbf{h},\mathbf{a})\right\rangle_{q(\mathbf{x},\mathbf{h})} &= \text{const} \quad (28)\\ &+ \frac{1}{2\sigma_n^2}\Big(\boldsymbol{\mu}_\mathbf{x}^T\big({\mathbf{K}^\mathbf{a}}^T\mathbf{R}\,\mathbf{K}^\mathbf{a} + {\mathbf{I}_\mathbf{y}}^T(\mathbb{I}-\mathbf{R})\,\mathbf{I}_\mathbf{y}\big)\,\boldsymbol{\mu}_\mathbf{x}\\ &\quad + \mathbf{r}^T(\mathbf{K}^\mathbf{a}\circ\mathbf{K}^\mathbf{a})\boldsymbol{\sigma}_\mathbf{x} + (\mathbf{1}-\mathbf{r})^T\mathbf{I}_\mathbf{y}\,\boldsymbol{\sigma}_\mathbf{x}\\ &\quad - 2\boldsymbol{\mu}_\mathbf{x}^T\big({\mathbf{K}^\mathbf{a}}^T\mathbf{R} + {\mathbf{I}_\mathbf{y}}^T(\mathbb{I}-\mathbf{R})\big)\mathbf{y} + \mathbf{y}^T\mathbf{y}\Big),\end{aligned}$$

where $\mathbf{R} \equiv \mathrm{diag}(\mathbf{r})$ and $\mathbb{I}$ is the identity matrix. Here, Eq. (28) extends the uniform case in [25] for non-uniform blur matrices $\mathbf{K}^\mathbf{a}$.

Inserting all expectations into Eq. (19), we obtain an explicit form of the free energy $F(q,\mathbf{a})$.

B. Update Equations for Stage 1

Next, we give the update equations w.r.t. the variational parameters of q as well as the motion parameters $\mathbf{a}$ during the first stage (in derivative space).

Latent indicator update. Levin *et al.* [18] have shown how to update the GSM (Gaussian scale mixture) indicators $q(\mathbf{l})$ in closed form. Adapting their derivation to our formulation, it is not difficult to see that

$$v_{i,\gamma,j} = \frac{1}{Z_{i,\gamma}}\exp\left(-\tfrac{1}{2\sigma_j^2}\hat{f}_{i,\gamma}\right)\frac{\pi_j}{\sigma_j}, \quad (29)$$

with

$$Z_{i,\gamma} = \sum_j \exp\left(-\tfrac{1}{2\sigma_j^2}\hat{f}_{i,\gamma}\right)\frac{\pi_j}{\sigma_j}, \quad (30)$$

$$\hat{f}_{i,\gamma} = \left(\mathbf{D}_\gamma\boldsymbol{\mu}_\mathbf{x}\circ\mathbf{D}_\gamma\boldsymbol{\mu}_\mathbf{x} + (\mathbf{D}_\gamma\circ\mathbf{D}_\gamma)\boldsymbol{\sigma}_\mathbf{x}\right)_i. \quad (31)$$

Image update. Isolating the terms involving $\boldsymbol{\mu}_\mathbf{x}$, we obtain the quadratic energy

$$F(q,\mathbf{a}) = \frac{1}{2}\boldsymbol{\mu}_\mathbf{x}^T\mathbf{A}_\mathbf{x}\boldsymbol{\mu}_\mathbf{x} + {\mathbf{b}_\mathbf{x}}^T\boldsymbol{\mu}_\mathbf{x} + \text{const}, \quad (32)$$

with

$$\begin{aligned}\mathbf{A}_\mathbf{x} &= \frac{1}{\sigma_n^2}\left({\mathbf{K}^\mathbf{a}}^T\mathbf{R}\,\mathbf{K}^\mathbf{a} + {\mathbf{I}_\mathbf{y}}^T(\mathbb{I}-\mathbf{R})\,\mathbf{I}_\mathbf{y}\right)\\ &\quad + \sum_{\gamma,j}\frac{1}{\sigma_j^2}\mathbf{D}_\gamma^T\,\mathrm{diag}(\mathbf{v}_{\gamma,j})\mathbf{D}_\gamma, \quad (33)\\ \mathbf{b}_\mathbf{x} &= -\frac{1}{\sigma_n^2}\left({\mathbf{K}^\mathbf{a}}^T\mathbf{R} + {\mathbf{I}_\mathbf{y}}^T(\mathbb{I}-\mathbf{R})\right)\mathbf{y}, \quad (34)\end{aligned}$$

where $\mathbf{v}_{\gamma,j}$ is a vector containing the parameters of the multinomial distribution in the j^{th} mixture component being associated with the γ^{th} derivative filter. Setting the gradient of Eq. (32) to zero yields a linear system, which can be solved efficiently, *e.g.* using conjugate gradient methods. We can apply similar steps to obtain the update equation for the diagonal covariance $\boldsymbol{\sigma}_\mathbf{x}$, which is given by the element-wise inverse of the diagonal of the linear system for $\boldsymbol{\mu}_\mathbf{x}$.

Segmentation update. As explained in the paper, we update the parameters of the Bernoulli distribution of the segmentation by variational message passing (Eq. 15). The required unary contributions are induced by both the bias term in the segmentation (Eq. 26) as well as the expected log-likelihood (Eq. 28):

$$\begin{aligned}\mathbf{g}(q(\mathbf{x}),\mathbf{a},\mathbf{y}) &= \lambda_0\mathbf{1} \quad (35)\\ &+ \frac{1}{2\sigma_n^2}\Big(\mathbf{K}^\mathbf{a}\boldsymbol{\mu}_\mathbf{x}\circ\mathbf{K}^\mathbf{a}\boldsymbol{\mu}_\mathbf{x} + (\mathbf{K}^\mathbf{a}\circ\mathbf{K}^\mathbf{a})\boldsymbol{\sigma}_\mathbf{x}\\ &\quad - 2\,\mathrm{diag}(\mathbf{K}^\mathbf{a}\boldsymbol{\mu}_\mathbf{x})\,\mathbf{y}\Big)\\ &- \frac{1}{2\sigma_n^2}\Big(\mathbf{I}_\mathbf{y}\boldsymbol{\mu}_\mathbf{x}\circ\mathbf{I}_\mathbf{y}\boldsymbol{\mu}_\mathbf{x} + \mathbf{I}_\mathbf{y}\,\boldsymbol{\sigma}_\mathbf{x} - 2\,\mathrm{diag}(\mathbf{I}_\mathbf{y}\boldsymbol{\mu}_\mathbf{x})\,\mathbf{y}\Big),\end{aligned}$$

where $\mathbf{1}$ is a vector of all ones.

Motion update. We now utilize the parametric nature of our model to efficiently minimize the free energy w.r.t. $\mathbf{a}$. To this end, note that the motion parameters $\mathbf{a}$ exclusively occur in the expected log-likelihood (Eq. 28) and essentially form a quadratic norm plus an additional term accounting for the uncertainty of the latent image. Unfortunately, the parameters $\mathbf{a}$ occur non-linearly within the blur matrix $\mathbf{K}^\mathbf{a}$, which makes it hard to obtain a closed-form solution. On the other hand, there are very efficient methods for minimizing non-linear least squares objectives, *i.e.* quadratic norms of non-linear residuals (see [40] for more details). Here, we will adapt such a highly efficient method to our formulation. To begin with, we observe that the motion parameters $\mathbf{a}$ exclusively occur in a subset of terms of the expected log-likelihood

$$\begin{aligned}F(q,\mathbf{a}) \propto \frac{1}{2}\Big(&\boldsymbol{\mu}_\mathbf{x}^T{\mathbf{K}^\mathbf{a}}^T\mathbf{R}\,\mathbf{K}^\mathbf{a}\boldsymbol{\mu}_\mathbf{x} + \mathbf{r}^T(\mathbf{K}^\mathbf{a}\circ\mathbf{K}^\mathbf{a})\boldsymbol{\sigma}_\mathbf{x}\\ &- 2\boldsymbol{\mu}_\mathbf{x}^T{\mathbf{K}^\mathbf{a}}^T\mathbf{R}\,\mathbf{y}\Big), \quad (36)\end{aligned}$$

where we have dropped constants as well as the factor σ_n^2, as they are not relevant for the minimization. We continue by linearizing the blur kernels $\mathbf{K}^{\mathbf{a}}$ around the operating point $\mathbf{a}_0$ and express Eq. (36) by means of the linearized blur kernels $\mathbf{K}^{\mathbf{d}}$ with the unknown increment vector $\mathbf{d} \equiv \mathbf{a} - \mathbf{a}_0$ (see Section 3 of the paper). Note that linear and quadratic terms involving the non-linear blur matrix can be easily approximated by terms that are linear in $\mathbf{d}$, *e.g.*

$$\mathbf{K}^{\mathbf{a}}\mathbf{x} \approx \mathbf{K}^{\mathbf{d}}\mathbf{x} = \mathbf{K}^0\mathbf{x} + \nabla_{\mathbf{a}}(\mathbf{K}^0\mathbf{x})\mathbf{d}, \quad (37)$$

$$\begin{aligned}\mathbf{r}^T(\mathbf{K}^{\mathbf{a}} \circ \mathbf{K}^{\mathbf{a}})\boldsymbol{\sigma}_{\mathbf{x}} &\approx \mathbf{r}^T(\mathbf{K}^{\mathbf{d}} \circ \mathbf{K}^{\mathbf{d}})\boldsymbol{\sigma}_{\mathbf{x}} \\ &= \mathbf{d}^T\mathbf{H}^0(\boldsymbol{\sigma}_{\mathbf{x}})\,\mathbf{d} + 2\,\mathbf{d}^T\mathbf{h}^0(\boldsymbol{\sigma}_{\mathbf{x}}) \\ &\quad + \mathbf{r}^T(\mathbf{K}^0 \circ \mathbf{K}^0)\boldsymbol{\sigma}_{\mathbf{x}}, \end{aligned} \quad (38)$$

where we define the $N \times 6$ matrix

$$\nabla_{\mathbf{a}}(\mathbf{K}^0\mathbf{x}) = \left(\frac{\partial \mathbf{K}^0}{\partial a_j}\mathbf{x}\right)_j, \quad (39)$$

the 6×6 matrix

$$\mathbf{H}^0(\boldsymbol{\sigma}_{\mathbf{x}}) = \left(\mathbf{r}^T\left(\frac{\partial \mathbf{K}^0}{\partial a_i} \circ \frac{\partial \mathbf{K}^{\mathbf{a}}}{\partial a_j}\right)\boldsymbol{\sigma}_{\mathbf{x}}\right)_{i,j}, \quad (40)$$

and the 6×1 vector

$$\mathbf{h}^0(\boldsymbol{\sigma}_{\mathbf{x}}) = \left(\mathbf{r}^T\left(\frac{\partial \mathbf{K}^0}{\partial a_i} \circ \mathbf{K}^0\right)\boldsymbol{\sigma}_{\mathbf{x}}\right)_i. \quad (41)$$

In the expressions above the advantage of the parametric model comes into play as we can efficiently compute $\nabla_{\mathbf{a}}(\mathbf{K}^0\mathbf{x})$, $\mathbf{H}^0(\boldsymbol{\sigma}_{\mathbf{x}})$ and $\mathbf{h}^0(\boldsymbol{\sigma}_{\mathbf{x}})$ by means of the derivative filters $\frac{\partial \mathbf{K}^0}{\partial a_i}$ (see Section 4 in the paper). Once we adopt this linearization, the free energy becomes (locally) quadratic in the unknown increment vector $\mathbf{d}$:

$$F(q, \mathbf{d}) = \frac{1}{2}\mathbf{d}^T\mathbf{A}_0\,\mathbf{d} + \mathbf{d}^T\mathbf{b}_0 + \text{const}, \quad (42)$$

with

$$\mathbf{A}_0 = \big(\nabla_{\mathbf{a}}(\mathbf{K}^0\boldsymbol{\mu}_{\mathbf{x}})\big)^T\mathbf{R}\,\nabla_{\mathbf{a}}\big(\mathbf{K}^0\boldsymbol{\mu}_{\mathbf{x}}\big) + \mathbf{H}^0(\boldsymbol{\sigma}_{\mathbf{x}}), \quad (43)$$

$$\begin{aligned}\mathbf{b}_0 = &\big(\nabla_{\mathbf{a}}(\mathbf{K}^0\boldsymbol{\mu}_{\mathbf{x}})\big)^T\mathbf{R}\,\mathbf{K}^0\boldsymbol{\mu}_{\mathbf{x}} + \mathbf{h}^0(\boldsymbol{\sigma}_{\mathbf{x}}) \\ &- \big(\nabla_{\mathbf{a}}(\mathbf{K}^0\boldsymbol{\mu}_{\mathbf{x}})\big)^T\mathbf{R}\,\mathbf{y}. \end{aligned} \quad (44)$$

We can now use this locally quadratic approximation to minimize the (non-linear) free energy around subsequent operating points. Furthermore, we can build upon regularization techniques from standard non-linear least squares methods, such as the Levenberg-Marquardt approach or use an even more sophisticated step-size control. In our implementation we rely on the Armijo rule [38].

C. Update Equations for Stage 2

Finally, we give the update equations for the variational parameters during the second stage (in image space).

Segmentation update. From Eq. (17) we obtain the free energy

$$\begin{aligned}\tilde{F}(q, \mathbf{a}) = &\; F(q, \mathbf{a}) - \Big\langle \lambda_c \sum_i h_i \log \text{GMM}(y_i \,|\, \theta_f) \quad (45)\\ &+ (1 - h_i)\log \text{GMM}(y_i \,|\, \theta_b)\Big\rangle_{q(\mathbf{h})} \\ = &\; F(q, \mathbf{a}) - \lambda_c \sum_i r_i \log \text{GMM}(y_i \,|\, \theta_f) \quad (46)\\ &+ (1 - r_i)\log \text{GMM}(y_i \,|\, \theta_b), \end{aligned}$$

augmented by one term accounting for the color statistics of the background/foreground, respectively. In turn, the update for the segmentation in stage 2 differs by one additional unary term:

$$\begin{aligned}\tilde{\mathbf{g}}(q(\mathbf{x}), \mathbf{a}, \mathbf{y}) = &\; \mathbf{g}(q(\mathbf{x}), \mathbf{a}, \mathbf{y}) \\ &+ \lambda_c\big(-\log \text{GMM}(y_i \,|\, \theta_f) \\ &+ \log \text{GMM}(y_i \,|\, \theta_b)\big)_i. \end{aligned} \quad (47)$$

Color statistics update. Let $\theta_f = \{\pi_{f,j}, \mu_{f,j}, \Sigma_{f,j} \,|\, j = 1 \ldots J\}$ and $\theta_b = \{\pi_{b,j}, \mu_{b,j}, \Sigma_{b,j} \,|\, j = 1 \ldots J\}$ be the parameters of the Gaussian mixture model for the foreground and background colors, respectively. Then Eq. (46) can be minimized by the expectation-maximization (EM) algorithm for Gaussian mixture models, however, each update equation is weighted by the parameters of the Bernoulli distribution of the segmentation.

For instance, updates for the foreground color statistics are given by

$$\alpha_{i,j} = \frac{\pi_j \mathcal{N}(\mathbf{y}_i \,|\, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \pi_k \mathcal{N}(\mathbf{y}_i \,|\, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}, \quad (48)$$

$$N_j = \sum_i r_i \alpha_{i,j}, \quad (49)$$

$$\boldsymbol{\mu}_j^{\text{new}} = \frac{1}{N_j}\sum_i r_i \alpha_{i,j}\mathbf{y}_i, \quad (50)$$

$$\boldsymbol{\Sigma}_j^{\text{new}} = \frac{1}{N_j}\sum_i r_i \alpha_{i,j}(\mathbf{y}_i - \boldsymbol{\mu}_j^{\text{new}})(\mathbf{y}_i - \boldsymbol{\mu}_j^{\text{new}})^T, \quad (51)$$

$$\pi_j^{\text{new}} = N_j / \sum_k N_k, \quad (52)$$

where we dropped the foreground index f for brevity.

Figure 8. Synthesized uniform and non-uniform motion blur.

D. Synthetic Dataset

For the quantitative analysis in the paper we created a dataset of 32 test images, divided in two subsets: uniform linear motion and non-uniform affine motion. The test images were created by extracting objects and segmentations from the VOC2012 dataset [39], and pasting them on top of different backgrounds. More precisely, motion blur is simulated by iteratively warping both the extracted images and segmentations according to either uniform or affine motion, and pasting its warped (latent) images on top of the background. While the resulting blurred image is given by the average of all warped latent images, the ground truth segmentation is chosen to be the maximum, *i.e*. the union, of all warped segmentations. Examples are shown in Fig. 8.

E. Sensitivity Analysis

We analyze the sensitivity of our approach to different initializations. To that end, we created a synthetic example (Fig. 8, left) of horizontal motion ($a_1 = 15$) and measured the resulting average endpoint error of the motion estimation for initializations with increasingly large motion in either vertical or horizontal direction. Table 2 shows the resulting average endpoint errors. Unless our method is initialized with a significant motion in the incorrect (vertical) direction, which leads to a failure (marked red), our algorithm yields consistent results. Initializing with a small motion avoids such issues.

Table 2. Average endpoint error after different initializations.

$\lvert a_{1/4} \rvert$	0.1	0.5	1	3	5	7
vertical	0.55	0.48	0.47	17.47	21.39	28.91
horizontal	0.53	0.56	0.51	0.57	0.52	0.53

F. Additional Results

We continue to show a few more results in addition to the ones provided in the main paper. Note that some of these examples are taken from [26].

Additional examples. While our approach is primarily aimed at recovering object (foreground) motion, the example in the last row of Fig. 7 of the main paper has already shown that we can also estimate motion and segmentation from a motion blurred background. Figures 9 and 10 show two more such examples in which a sharp bicyclist is shown before a motion-blurred background. Our approach correctly identities the background scene as the motion-blurred region ("foreground") and vice versa.

Figures 11 and 12 show additional results for a purely rotational ferris wheel, as well as a motion-blurred rollercoaster. Note that the rollercoaster is segmented very well, but the ferris wheel less so. While our variational framework identities the rotational motion correctly, the blurry foreground of the outer wheel blends with the background, hence our approach does not properly pick up these regions as part of the blurry foreground.

Failure cases. In Figs. 13 and 14 we show two examples for which the variational framework fails to estimate either a correct motion model or a segmentation. In the first example (Fig. 13) both estimating the motion model as well as estimating the segmentation fails. To be successful, our approach requires a sufficiently large region of observable motion blur; if this is not the case our algorithm may end up in poor local minima. This is in particular the case for estimating the correct motion, since the motion updates are based on iteratively optimizing a non-linear objective by (locally) quadratic approximations. Also observe how the inference heavily picks up the horizontal structures in the background wall on the right-hand side, as they provide evidence for horizontal motion blur.

The second example (Fig. 14) shows how the variational framework fails to estimate the correct motion model due to directional ambiguities. Note how the top left motion vectors point to the left, while a major part of the estimated motion vectors point to the right. In our approach we tackle this ambiguity by explicitly modeling symmetric blur kernels, however, the motion parameters still allow for two equally good explanations: Either the background translates to the left or to the right. In practice we overcome this problem by initializing the motion estimate with a slight bias towards either direction. However, this example indicates that this bias alone may not always be enough to resolve the ambiguities one may observe during the inference process.

References

[38] L. Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific J. Math.*, 16(1):1–3, 1966. 3

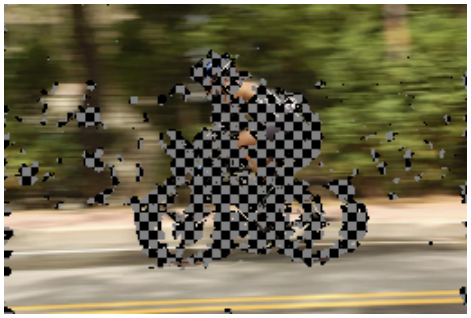
[39] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results, 2012. 4

[40] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, NY, USA, 2nd edition, 2006. 2

[41] K. B. Petersen and M. S. Pedersen. The matrix cookbook, 2012. Version 20121115. 2

(a) Blurry input

(b) Parametric motion + motion segmentation

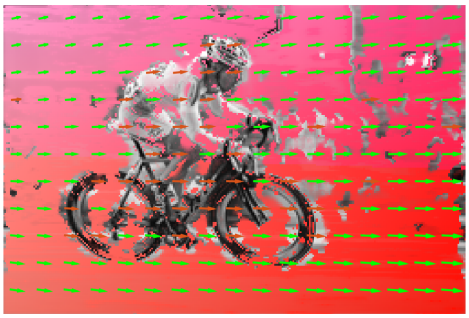


(c) Blurry foreground

(d) Static background

Figure 9. Motion from a blurry background.

(a) Blurry input

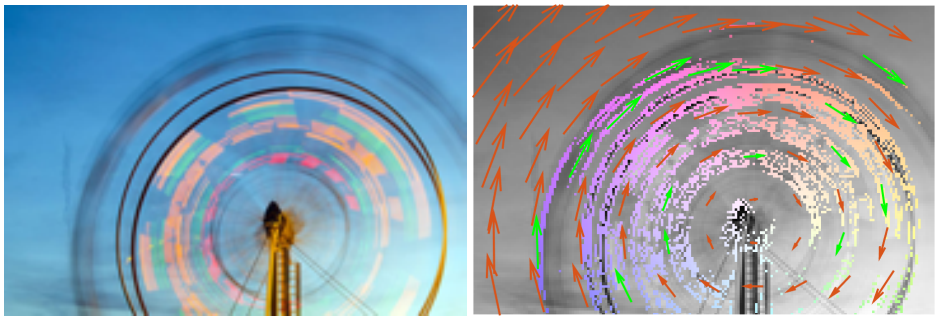


(b) Parametric motion + motion segmentation

(c) Blurry foreground

(d) Static background

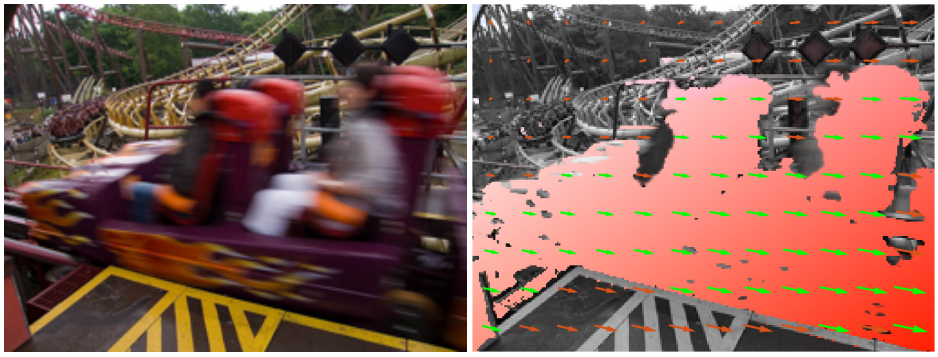
Figure 10. Motion from a blurry background.



(a) Blurry input (b) Parametric motion + motion segmentation

(c) Blurry foreground (d) Static background

Figure 11. Rotational motion from a ferris wheel.



(a) Blurry input (b) Parametric motion + motion segmentation

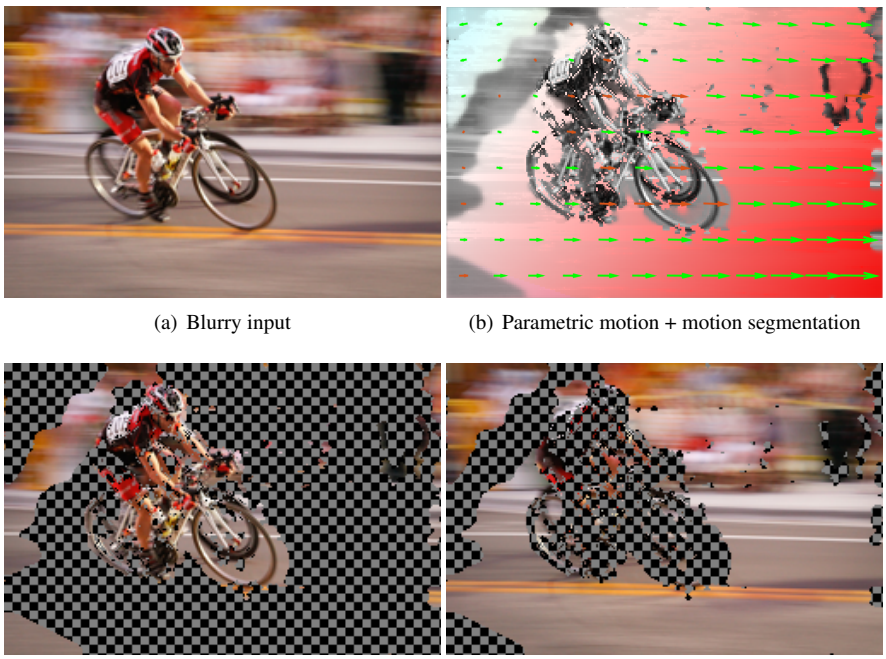
(c) Blurry foreground (d) Static background

Figure 12. Affine motion from a rollercoaster.

(a) Blurry input (b) Parametric motion + motion segmentation

(c) Blurry foreground (d) Static background

Figure 13. Motion estimation may fail if the blurry region is too small in comparison to the background.



(a) Blurry input (b) Parametric motion + motion segmentation

(c) Blurry foreground (d) Static background

Figure 14. Motion estimation may fail due to directional ambiguities.