Shape from Blur: Recovering Textured 3D Shape and Motion of Fast Moving Objects
– Supplementary Material –

Denys Rozumnyi
Department of Computer Science
ETH Zurich, Switzerland
denys.rozumnyi@inf.ethz.ch

Martin R. Oswald[1,2]
[2]University of Amsterdam
Netherlands
martin.oswald@inf.ethz.ch

Vittorio Ferrari
Google Research
Zurich, Switzerland
vittoferrari@google.com

Marc Pollefeys
[1]Department of Computer Science
ETH Zurich, Switzerland
marc.pollefeys@inf.ethz.ch

A Loss terms

This section provides additional details about the formulation of the loss terms.

Texture smoothness loss. The texture smoothness loss $\mathcal{L}_T(\Theta)$ is implemented by total variation. If the number of pixels is K, and the derivative of the estimated texture map is T_x in x image direction and T_y in y image direction, then the loss is defined as

$$\mathcal{L}_T(\Theta) = \frac{1}{K} \sum_{p=1}^{N} |T_x(p)| + |T_y(p)| \ . \tag{1}$$

Laplacian shape regularization loss. To define the Laplacian shape regularization loss $\mathcal{L}_L(\Theta)$, we follow the notation proposed in the Soft Rasterizer method [1]. We assume that the mesh Θ consists of a set of vertices, where each vertex v_i is a 3-dimensional vector. Then, the function $N(v_i)$ represents the neighbors of vertex v_i: a set of all adjacent vertices as defined by the faces in Θ. The loss is the sum of differences between each vertex and the center of mass of all its neighbors:

$$\mathcal{L}_L(\Theta) = \sum_i \|v_i - \frac{1}{N(v_i)} \sum_{j \in N(v_i)} v_j\|_2^2 \ . \tag{2}$$

Intersection over Union. We clarify the definition of the Intersection over Union (IoU), which we use in the silhouette consistency loss in the main paper. The IoU between two masks (real-valued) or silhouettes (binary) is defined as

$$\text{IoU}(M_1, M_2) = \frac{\|M_1 \cdot M_2\|_1}{\|M_1 + M_2 - M_1 \cdot M_2\|_1} \ . \tag{3}$$

Losses in the ablation study. In the ablation study, we experimented with an alternative form of the silhouette consistency loss, which is the difference image-based loss $\mathcal{L}_D$. It uses simple background subtraction instead of sub-frame DeFMO [2] masks. The difference image is binarized as $D = \|I - B\|_2 > 0.1$. The loss is implemented as the intersection over union (IoU) between D and the estimated silhouettes, averaged over the exposure duration, and binarized:

$$\mathcal{L}_D(\Theta, \mathbf{r}, \Delta\mathbf{r}, \mathbf{t}, \Delta\mathbf{t}) = 1 - \text{IoU}\Big(D, \int_0^1 \mathcal{R}_S\big(\mathcal{M}(\Theta, \mathbf{r} + \tau \cdot \Delta\mathbf{r}, \mathbf{t} + \tau \cdot \Delta\mathbf{t})\big)\, \mathrm{d}\tau > 0\Big) \ . \tag{4}$$

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

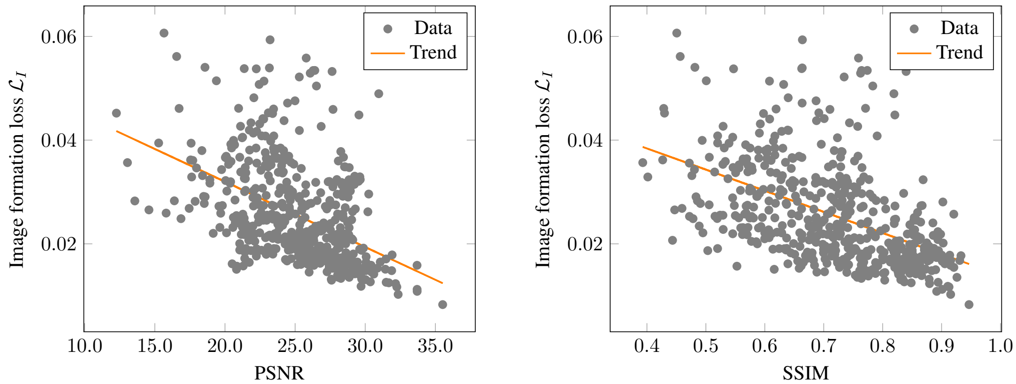

Figure 1: **Correlation between the image formation loss and evaluation metrics (PSNR/SSIM).** Linear regression of the measured data shows the trend: the lower is the image formation loss, the better are the evaluation metrics. Thus, it can be viewed as a surrogate for the evaluation metrics.

We also experimented with another loss $\mathcal{L}_F$, favoring the textures rendered by the proposed method to be similar to the sub-frame appearances estimated by DeFMO. The mathematical form is similar to the silhouette consistency loss, but instead of the masks, it directly compares the RGB values:

$$\mathcal{L}_F(\Theta, \mathbf{r}, \Delta\mathbf{r}, \mathbf{t}, \Delta\mathbf{t}) = \int_0^1 \|F_\tau - \mathcal{R}_F\big(\mathcal{M}(\Theta, \mathbf{r} + \tau \cdot \Delta\mathbf{r}, \mathbf{t} + \tau \cdot \Delta\mathbf{t})\big)\|_1 \, \mathrm{d}\tau \ . \quad (5)$$

B Correlation between image formation loss and evaluation metrics

As mentioned in the main paper, we noticed that the image formation loss is directly related to the quality of reconstruction and deblurring as measured by the official evaluation metrics PSNR and SSIM on the FMO deblurring benchmark [2]. Fig. 1 shows all measured data over all input image sequences in all three datasets from the FMO benchmark. The orange line demonstrates the observed linear trend.

C Additional qualitative results

We show additional results in Fig. 2 for the TbD-3D dataset [3] and in Fig. 3 for the Falling Objects dataset [4].

References

[1] S. Liu, T. Li, W. Chen, and H. Li, “Soft rasterizer: A differentiable renderer for image-based 3d reasoning,” *ICCV*, Oct 2019.

[2] D. Rozumnyi, M. R. Oswald, V. Ferrari, J. Matas, and M. Pollefeys, “Defmo: Deblurring and shape recovery of fast moving objects,” in *CVPR*, (Nashville, Tennessee, USA), Jun 2021.

[3] D. Rozumnyi, J. Kotera, F. Šroubek, and J. Matas, “Sub-frame appearance and 6d pose estimation of fast moving objects,” in *CVPR*, pp. 6777–6785, 2020.

[4] J. Kotera, J. Matas, and F. Šroubek, “Restoration of fast moving objects,” *IEEE TIP*, vol. 29, pp. 8577–8589, 2020.

[5] J. Kotera, D. Rozumnyi, F. Šroubek, and J. Matas, “Intra-frame object tracking by deblatting,” in *ICCVW*, Oct 2019.

[6] M. Jin, G. Meishvili, and P. Favaro, “Learning to extract a video sequence from a single motion-blurred image,” in *CVPR*, June 2018.

[7] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, “Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better,” in *ICCV*, Oct 2019.

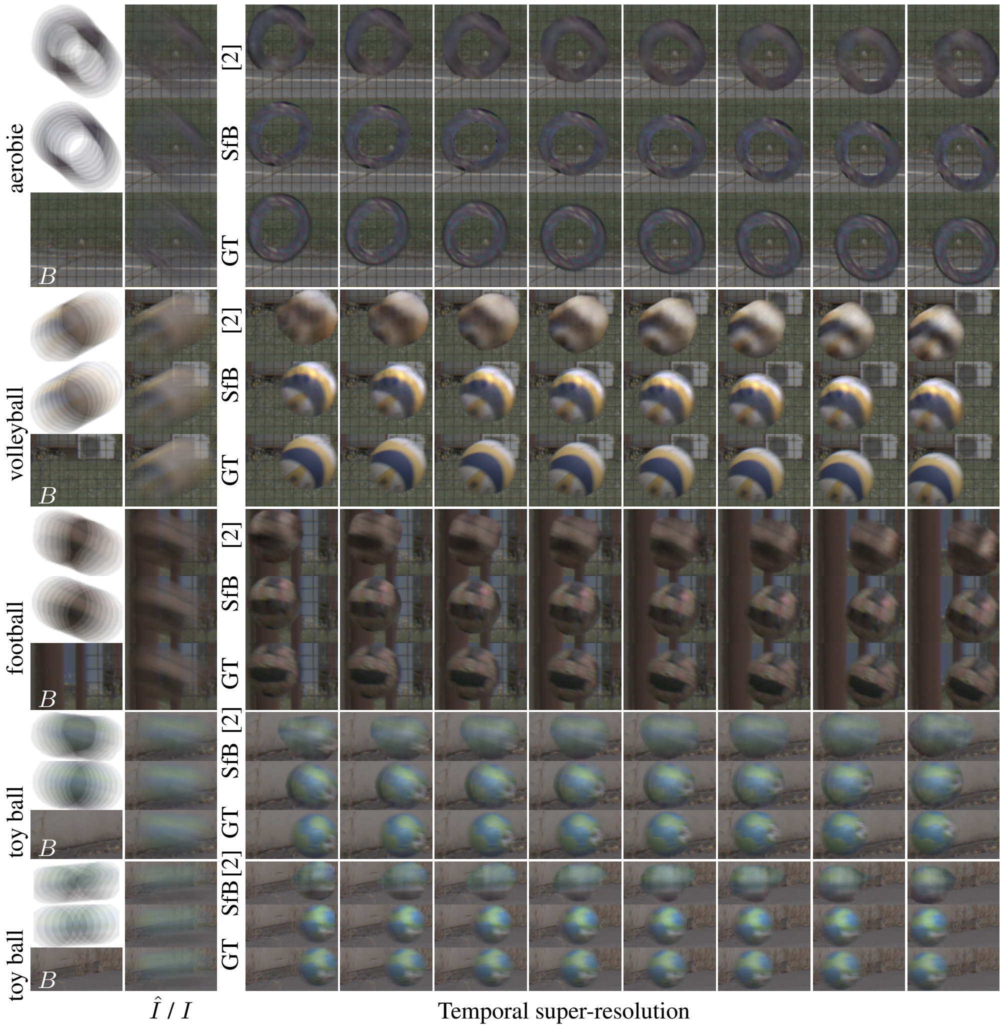

Figure 2: **Additional results on the TbD-3D [3] dataset.** We compare the proposed Shape-from-Blur (SfB) method with the previous state-of-the-art DeFMO [2], and also show the ground truth from a high-speed camera (GT). The actual input image I is almost indistinguishable from the image $\hat{I}$ rendered by SfB as a mixture of the background B and the reconstructed object appearance, temporally averaged over the image exposure time (shown above the background).

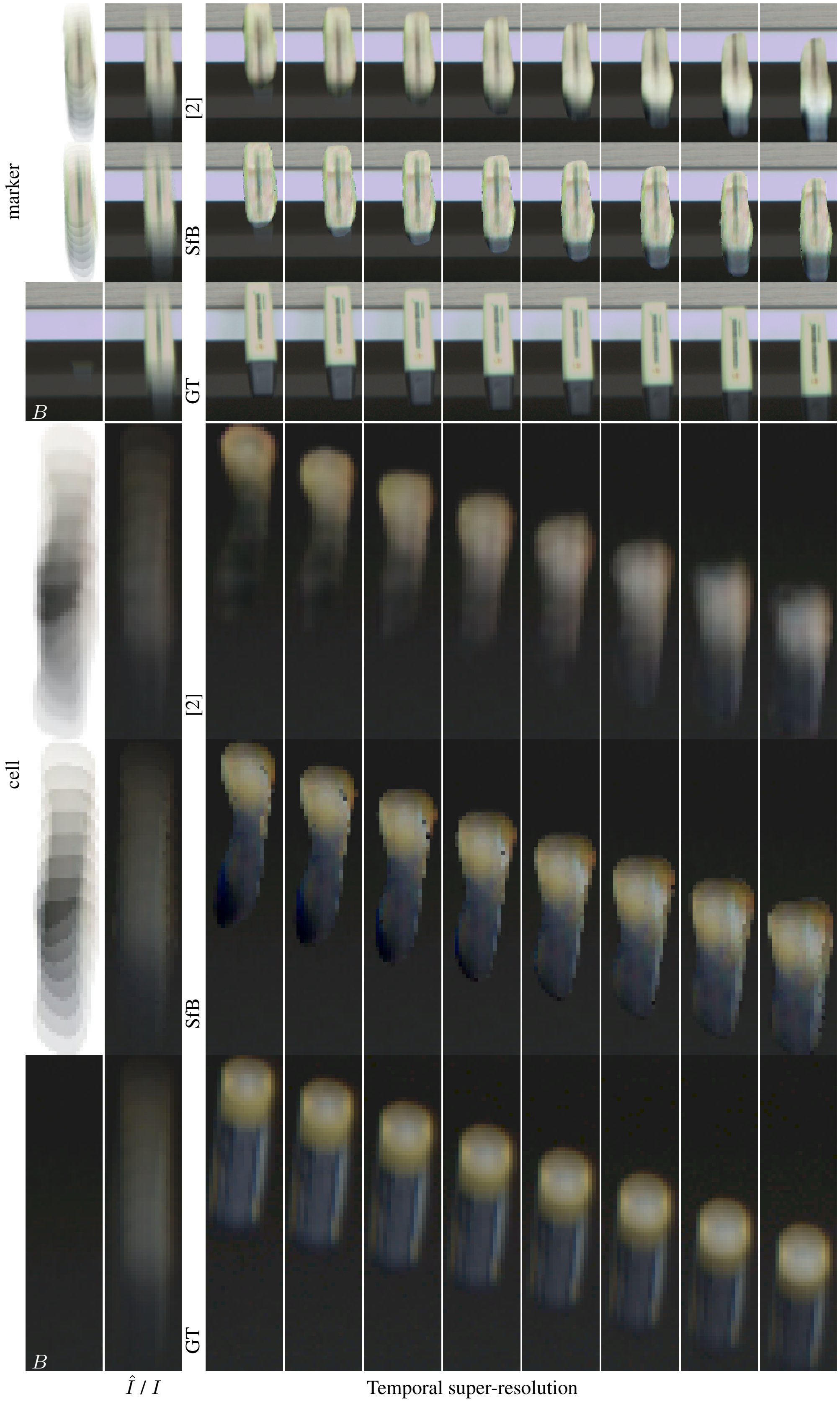

Figure 3: **Additional results on the Falling Objects [4] dataset.**

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Limitations are discussed in Sec. 5.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] We believe there are no direct negative societal impacts of our work.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A] We do not include theoretical results.
   (b) Did you include complete proofs of all theoretical results? [N/A] We do not include theoretical results.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] We include the code to reproduce the main experimental results in the supplemental material. The full code will be publicly available after acceptance.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] The training details are provided in Sec. 3.2.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] We do not report error bars since the proposed method is deterministic and is not based on random number generator. The initialization of the proposed optimization is always the same as reported in Sec. 3.2.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] The total amount of compute and used resources are reported in Sec. 3.2.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We cited the creators of DeFMO [2], which we use directly in the proposed method. Also, the creators of all used datasets are cited [5, 3, 4]. The other compared methods are also cited [6, 7]. The used datasets are available at `http://cmp.felk.cvut.cz/fmo/`. The FMO benchmark implementation is taken from `https://github.com/rozumden/fmo-deblurring-benchmark`. DeFMO implementation is taken from `https://github.com/rozumden/DeFMO`.
   (b) Did you mention the license of the assets? [Yes] The mentioned assets are open-sourced.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] Implementation of the proposed method is included in the supplemental material.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] The used data is open-sourced.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] The used data does not contain personally identifiable information or offensive content.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]