

Development of a Camera Motion Estimation Method Utilizing Motion Blur in Images

Yuxin Zhao^[0009-0007-6472-9588], Hirotake Ishii^[0000-0002-5638-4862], and
Hiroshi Shimoda^[0000-0002-5231-4955]

Graduate School of Energy Science, Kyoto University, Kyoto, Japan
{zhao,hirotake,shimoda}@ei.energy.kyoto-u.ac.jp

Abstract. Motion blur presents a significant challenge in visual SLAM. Since a satisfying performance highly relies on clear and feature-rich images, the system will easily fail to extract features and lose track due to motion blur. To address this issue, this study introduces a novel solution that estimates the camera motion using blur features extracted from images. The proposed approach models the relationship between camera motion and motion blur, creating a comprehensive motion blur dataset labeled with camera motion. By decoupling the motion estimation process into predicting magnitude and direction separately, a neural network is trained using blur and depth images as inputs to output the camera motion. Experimental results demonstrate that the proposed method successfully estimates motion from blur, even in long-term blur scenarios. This method can potentially serve as an auxiliary motion estimation module to enhance the robustness and accuracy of the visual odometer when motion blur is encountered.

Keywords: Motion Blur · Camera Motion Estimation · Deep Learning.

1 Introduction

Recovering the 3D scene structure and camera motion from images is crucial in various computer vision tasks, such as 3D reconstruction, robot navigation, and virtual/augmented reality [1]. Traditionally, the classical approach involves extracting and matching visual features in the images. Geometric constraints and optimization algorithms are then employed to estimate the camera pose. However, this process heavily relies on clear and distinctive features within the images [2]. Consequently, classical algorithms often encounter significant errors or even fail to function altogether in scenarios with visual degradation. One common form of visual degradation is motion blur, which occurs when the camera is exposed for a long period or experiences rapid movement [3]. Motion blur poses a significant challenge as it hampers the extraction of sufficient features from the images.

In recent studies, researchers have made efforts to address this issue. One approach involves incorporating line features in addition to point features, as they offer a more robust representation of the scene's geometry [4]. By doing so,

these researchers were able to achieve improved accuracy and efficiency in estimating camera pose under motion blur scenarios, surpassing the performance of classical algorithms. Despite the advancements made, a limitation remains when the system encounters long-term abrupt camera motion [5]. In such cases, the tracking system still fails to perform adequately. Instead of proposing a comprehensive framework, some researchers focused on removing motion blur and estimating the camera pose from the restored images [6]. In another approach, researchers attempted to reblur a clear reference image and subsequently trained or optimized it to obtain the camera motion [7, 8]. Unfortunately, this technique relies on the availability of a reference image, which may not always be present in real-world application scenarios.

This research aims at estimating camera motion under long-term motion blur without relying on clear images. Since the appearance of motion blur depends on camera motion, it still retains valuable information about the motion [9]. Therefore, instead of treating motion blur as a visual degradation to be eliminated, we propose leveraging it to directly estimate camera motion through deep learning. The network learns blur features from motion blur image and output the relative camera pose during the formation process. Hopefully, the proposed method can serve as an auxiliary motion estimation module when the motion blur appears to improve the robustness and accuracy of the visual odometer.

2 Modeling of Blur and Motion

Training a neural network often requires a significant amount of datasets. However, datasets that include motion blur images with varying camera poses are scarce. Hence, as the basis of creating datasets, we introduced Q-Slerp interpolation algorithm into blur formation model in [7] and modeled the process of camera motion forming motion blur.

A digital camera forms an image by collecting the light during exposure time and converting it into electrical signals. When the camera is exposed for a long time or moving quickly, there will be a noticeable blur of drag marks in the image, known as motion blur. This process can be modeled as integrating over a set of virtual sharp images.

$$B = \frac{1}{\tau} \int_0^{\tau} C_t dt \quad (1)$$

where C_t indicates the virtual sharp image captured at time t during the total exposure time τ , and B is the final blur image. It can be approximated by averaging n virtual images C_i in the process:

$$B \approx \frac{1}{n} \sum_{i=0}^{n-1} C_i \quad (2)$$

The camera pose T is represented by the combination of translation vector and Euler angle, which contains 6 degrees of freedom (DOF):

$$T = [t_x, t_y, t_z, r_x, r_y, r_z] \quad (3)$$

where t_i , r_i ($i = x, y, z$) denote the translation and rotation along x, y, z axis respectively.

Given the virtual sharp image C_0 and depth image D_0 captured when $i = 0$, the total movement T during exposure time τ can be interpolated as n mid-poses $S_i(T)$. Each virtual image C_i can be modeled as the image transformed from C_0 with the interpolated pose $S_i(T)$:

$$C_i = \pi \left(S_i(T) \cdot \pi^{-1} (C_0, D_0) \right) \quad (4)$$

where π is the camera projection function which projects 3D point cloud into 2D image plane, π^{-1} is the inverse projection function which transform 2D point in image plane with depth into 3D point cloud.

3 Dataset Creation

After building the blur formation model, we utilized the proposed model and created a large dataset using computer graphics (CG) techniques.

First, the clear image C_0 and depth image D_0 pairs are captured through simulation based on HQ Suburban House [10] on Unity. Next, the relative poses T are generated randomly, and Q-Slerp algorithm is utilized to interpolate n mid-poses $S_i(T)$ among each relative pose. Then, by projecting clear and depth image pairs into 3D point cloud, the transformation of each mid-pose is operated on point cloud. Finally, the virtual sharp image C_i is generated by projecting the transformed point cloud onto image plane, and the blur image B is synthesized by taking the average of all images.

Through continuous testing, n was set as 64, which can properly simulate the motion blur. The rotation range was randomly generated around $[-0.05, 0.05]$ radians, while the translation range was around $[-0.25, 0.25]$ meters. During training, the translation was regularized to $[-0.05, 0.05]$ to maintain consistency across the results. Totally 15,632 images featuring motion blur caused by random motion on t_x and t_y are generated in this dataset. The sample of generated images is shown in Figure 1.

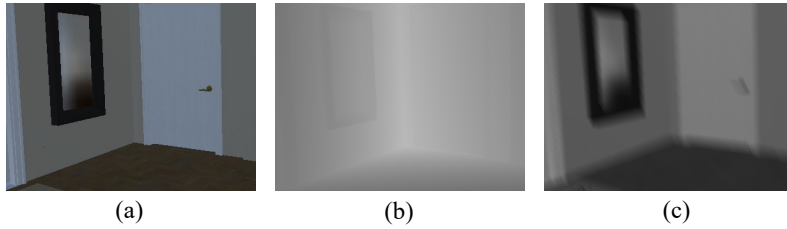


Fig. 1. Sample of dataset. (a) clear image; (b) depth image; (c) blur image.

4 Network Design and Training

4.1 Network Design

The network was designed to input blur and depth images and output camera motion. Estimating camera motion in a single step is challenging for the network to learn relevant features. Therefore, this process is decoupled to predict motion magnitude and direction separately. The diagram is illustrated in Figure 2. The ResNet-50 and ResNet-34 [11] are used as the backbone of magnitude module and direction module respectively. Two fully connected layers are added after the ResNet. The resolution of the input images is 640×480 .

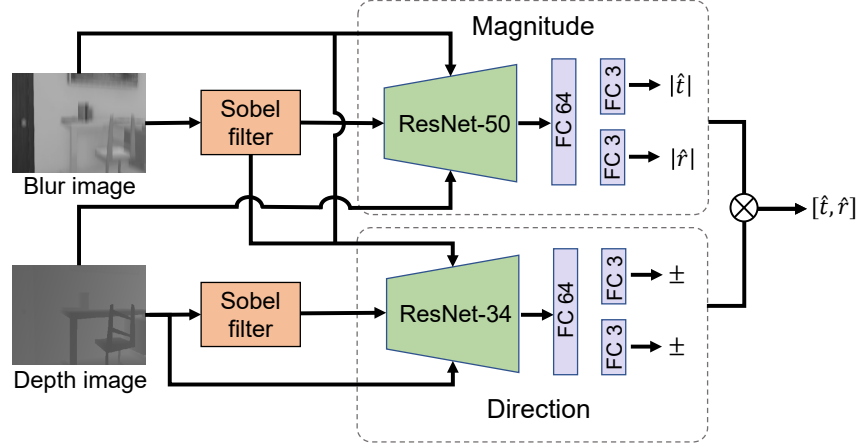


Fig. 2. Network architecture.

Magnitude Module This module is intended to estimate the magnitude of camera motion on each DOF from the blur amount in the image. Monocular images inherently lack scale information, making it impossible to restore motion scale using solely blur image. Hence, we use depth images to help predict actual camera motion. To prompt the neural network to extract blur information, the Sobel filter is utilized on blur images to emphasize the amount of blur. The resultant output of blur image is fed into the magnitude module together with the blur and depth image. The network outputs the absolute value of the translation vector $|\hat{t}|$ and rotation angle $|\hat{r}|$.

Direction Module The direction module classifies the positive and negative motion on each DOF from blur and depth images. When the camera is in motion, the shutter speed and timing for opening the shutter differ between capturing the

color and depth images. As a result, the areas shown in color and depth images exhibit slight difference depending on the camera motion direction. Leveraging this difference, we can estimate the direction of camera motion. To emphasize the disparity information between blur and depth image, the Sobel filter is applied on both of them, and taken as an input with the blur and depth image. The network would judge whether the direction is positive or negative. Finally, the outputs of two modules are combined to estimate the translation vector $\hat{t}$ and rotation vector $\hat{r}$.

4.2 Network Training

We trained the network on the dataset generated in Section 3, utilizing the Adam optimizer with a learning rate of 10^{-4} during the training process. The number of epochs was set to 300. For the magnitude module, the loss function employed was Root Mean Squared Error (RMSE) loss, calculated based on the absolute value of the pose. As for the direction module, Cross Entropy Loss was utilized for the classification of positive and negative.

5 Experiments

To evaluate the performance of the proposed method, we conducted test on datasets with long-term motion blur, where each image exhibited varying degrees of motion blur. Specifically, the datasets featured motion blur of translation along x-axis and y-axis movement were synthesized through CG in a different scene from the training datasets. Originating from (0, 0), the trajectories traced a counterclockwise heart-shaped pattern with intervals of 50mm, 100mm, 150mm, 200mm, and 250mm, indicating increasing degree of blur. This is a challenging task because no clear reference is available for the system to extract features.

We applied the networks containing magnitude and direction modules on the datasets. On average, 92.45% of the samples were correctly classified in the correct direction. Combining with the magnitude estimation, the estimated trajectories of the camera movement are indicated in Figure 3 in comparison with the ground truth. The error of each position is presented in grey lines. For the quantitative evaluation, absolute trajectory error (ATE) and the relative error (RE) were calculated in Table 1. The ATE of each case was to demonstrate the RMSE of ground truth pose and estimated pose, while the RE was calculated as scaling the RMSE by the range of the heart-shaped trajectory.

The results indicate that the estimated trajectory exhibits a consistency with the ground truth, especially in the case of 100mm. Notably, the presence of insufficient (50mm) motion blur adversely affects the performance. However, in the practical application, less motion blur generally represents clearer images, allowing traditional SLAM algorithms to function effectively. Therefore, this error is not influential. In addition, the accumulation of errors in frame-to-frame estimation leads to drift over time, which is more pronounced on the right side of the heart-shaped trajectory, resulting in a deviation from the ground truth. This issue can be well mitigated by loop closure detection and optimization.

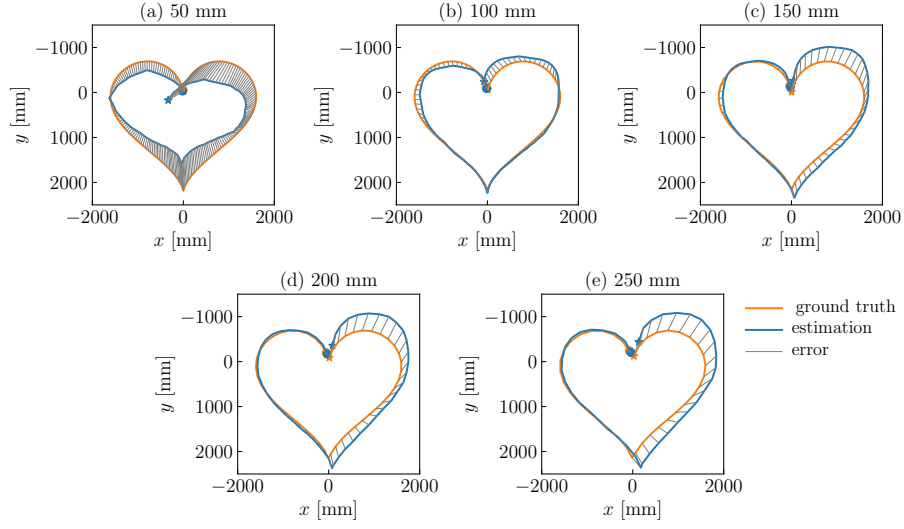


Fig. 3. Ground truth, estimation, and error of the test dataset. Circle and star denote the start and end points.

Table 1. Quantitative evaluation

Case	ATE [mm]	RE[%]
50 mm	339.94	11.58
100 mm	113.08	3.80
150 mm	197.75	6.85
200 mm	225.38	7.74
250 mm	264.34	8.95

6 Conclusion and Future Work

In conclusion, the proposed method shows promise for estimating camera motion from motion blur. Note that we achieved promising performance on the severe long-term blurred datasets, which is a challenging scenario for existing methods. It shows potential not only in estimating camera motion in the absence of inertial sensors but also in providing more precise estimates than inertial sensors. Future work includes expanding the model to accommodate more complex camera motions without sacrificing accuracy. Additionally, since only synthesized images are trained and evaluated in this study, the proposed approach will be generalized and evaluated on real-world datasets.

References

1. Fang, B., Zhan, Z.: A visual slam method based on point-line fusion in weak-matching scene. *International Journal of Advanced Robotic Systems* 17(2), 1729881420904193 (2020)
2. Bujanca, M., Shi, X., Spear, M., Zhao, P., Lennox, B., Luján, M.: Robust slam systems: Are we there yet? In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 5320–5327. IEEE (2021)
3. Carbajal, G., Vitoria, P., Delbracio, M., Musé, P., Lezama, J.: Non-uniform blur kernel estimation via adaptive basis decomposition. *arXiv preprint arXiv:2102.01026* (2021)
4. Li, D., Liu, S., Xiang, W., Tan, Q., Yuan, K., Zhang, Z., Hu, Y.: A slam system based on rgbd image and point-line feature. *IEEE Access* 9, 9012–9025 (2021)
5. Wang, Q., Yan, Z., Wang, J., Xue, F., Ma, W., Zha, H.: Line flow based slam. *arXiv preprint arXiv:2009.09972* (2020)
6. Lee, H.S., Kwon, J., Lee, K.M.: Simultaneous localization, mapping and deblurring. In: 2011 International Conference on Computer Vision. pp. 1203–1210. IEEE (2011)
7. Liu, P., Zuo, X., Larsson, V., Pollefeys, M.: Mba-vo: Motion blur aware visual odometry. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 5550–5559 (2021)
8. Nimisha, T., Rengarajan, V., Ambasamudram, R.: Semi-supervised learning of camera motion from a blurred image. In: 2018 25th IEEE International Conference on Image Processing (ICIP). pp. 803–807. IEEE (2018)
9. Argaw, D.M., Kim, J., Rameau, F., Cho, J.W., Kweon, I.S.: Optical flow estimation from a single motion-blurred image. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 35, pp. 891–900 (2021)
10. HQ Suburban House, <https://assetstore.unity.com/packages/3d/environments/urban/hq-suburban-house-81890>.
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)