Supplementary Material: Acquiring a Dynamic Light Field through a Single-Shot Coded Image

Ryoya Mizuno[†], Keita Takahashi[†], Michitaka Yoshida[‡], Chihiro Tsutake[†], Toshiaki Fujii[†], Hajime Nagahara[‡]
[†]Nagoya University, Japan, [‡]Osaka University, Japan

A. Light-Field Dataset and Motion Patterns

We summarize the light-field dataset and virtual motion patterns in Figs. A.1 and A.2. For the number of views, we mainly mention the case with 5×5 views in the main paper, where we extracted 223,020 patches from 51 light fields. We also considered another case with 8×8 views, where we extracted 128,520 light-field patches from 30 light fields. For the motion patterns assumed to create pseudo dynamic training samples, we mainly mention the case with $\alpha_x, \alpha_y \in \{-2, -1, 0, 1, 2\}$ $(16 + 8 + 1$ patterns) in the main paper, but we also considered other cases with reduced motions: $\alpha_x, \alpha_y \in \{-1, 0, 1, \}$ $(8 + 1$ patterns) and no motions.

5×5 views (51 light fields)
Chess, Lego Bulldozer, Lego Truck, Eucalyptus Flowers, Amethyst, Bracelet, The Stanford Bunny, Jelly Beans, Lego Knights, Tarot Cards and Crystal Ball (small angular extent), Treasure Chest (Stanford [1]), Red Dragon, Happy Buddha, Messerschmitt, Dice, Green Dragon, Mini Cooper, Butterfly, Lucy (MIT [4]), Bedroom, Bicycle, Herbs, Origami, Boxes, Cotton, Sideboard, Antinous, Boardgames, Dishes, Greek, Museum, Pens, Pillows, Platonic, Rosemary, Table, Tomb, Town, Vinyl (New HCI [3]), Buddha, Buddha 2, StillLife, Papillon, MonaRoom, Medieval, Horse, Couple, Cube, Maria, Pyramid, Statue (Old HCI [2])

8×8 views (30 light fields)
Chess, Lego Bulldozer, Lego Truck, Eucalyptus Flowers, Amethyst, Bracelet, The Stanford Bunny, Jelly Beans, Lego Knights, Tarot Cards and Crystal Ball (small angular extent), Treasure Chest Bedroom (Stanford [1]), Bicycle, Herbs, Origami, Boxes, Cotton, Sideboard, Antinous, Boardgames, Dishes, Greek, Museum, Pens, Pillows, Platonic, Rosemary, Table, Tomb, Town, Vinyl (New HCI [3])

Figure A.1. Light-field dataset used for training.

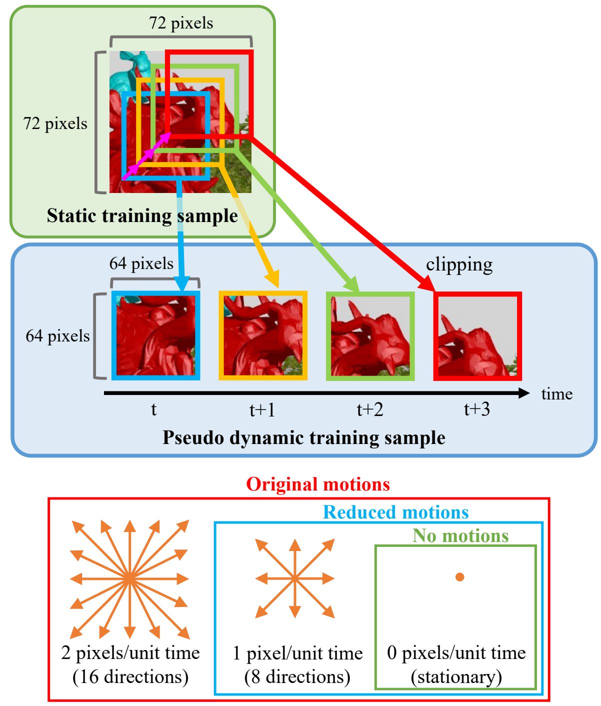


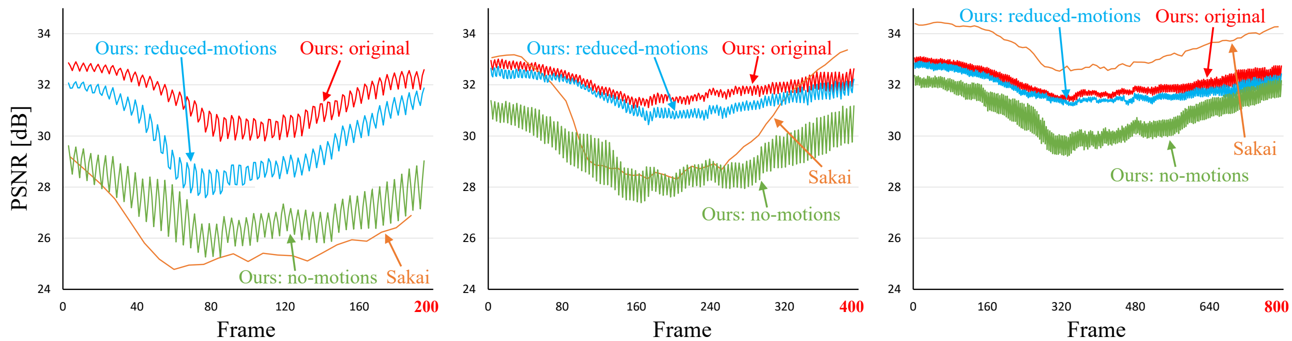
Figure A.2. Virtual motions given to the training samples.

B. Additional Evaluations

B.1. Speed of Scene Motions

We trained our network on the 5×5-view dataset but with two additional motion configurations: the reduced motions $(\alpha_x, \alpha_y \in \{-1, 0, 1, \})$ and no motions shown in Fig. A.2. We compared their performances against the original version $(\alpha_x, \alpha_y \in \{-2, -1, 0, 1, 2\})$. For evaluation, we used *Planets* scene, but changed the speed of the scene motions in several different levels: the original speed, $\times 0.5$ speed, and $\times 0.25$ speed. As shown in Fig. B.1, the reduced-motion and no-motion versions achieved high reconstruction quality at $\times 0.25$ speed. However, they suffered severe quality degradation as the speed of scene motions increased to the original speed. Meanwhile, the original version retained high reconstruction quality even for the original speed.

In Fig. B.1, we also present the performance of Sakai



(a) Original speed

(b) $\times 0.5$ speed

(c) $\times 0.25$ speed

Figure B.1. Speed of scene motion and reconstruction quality over time. From left to right, the scene motions were at original speed, $\times 0.5$ speed, and $\times 0.25$ speed. Our method was trained with different motion configurations.

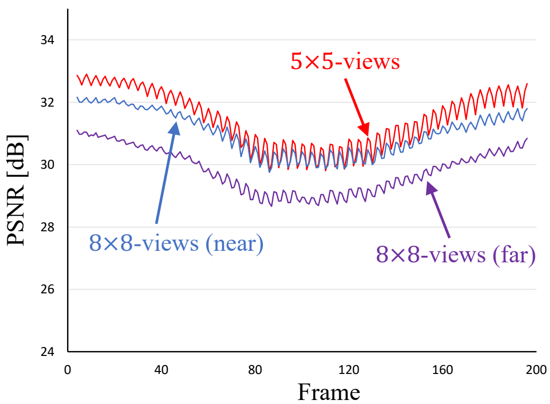


Figure B.2. Quantitative reconstruction quality along time for 5×5 views and 8×8 views test scenes.

et al. [5], because it largely depended on the speed of scene motions. When the scene motions were very slow (at $\times$ 0.25 speed in (c)), Sakai et al.'s method performed better than our method. This can be explained by the different ratios between the observed and reconstructed data; Sakai et al.'s method reconstructed a light field for a single time unit from three observed images (with the ratio of 3/25), while our method reconstructed a light field over four time units from a single observed image (with the ratio of 1/100). However, Sakai et al.'s method suffered drastic quality degradation as the motion speed increased. This degradation can be attributed to the assumption on the scene motions; they assumed relatively small motions (at most $\pm$2 pixels/frame) when creating the training dataset. Meanwhile, our method can accommodate a larger amount of motions. In fact, we assumed the same amount of motions (at most $\pm$2 pixels per unit time) with respect to the unit time. However, in our case, four unit times were included in a single exposure time of the camera. Therefore, in our case, the speed assumed for scene motions was at most $\pm$8 pixels/frame in terms of the camera's frame-rate, which led to better adaptability for faster scene motions.

B.2. 5×5 vs. 8×8 viewpoints

We trained our network on the 8×8-viewpoint dataset (with minimum modification to handle different numbers of views) and compared the performance with the 5×5-view version mentioned in the main paper. We used *Planets* scene for evaluation. For the case with 8×8 views, we tested two configurations. In the first configuration (8×8 views (near)), the viewpoint intervals were reduced to 4/7 times those for the 5×5 views, which kept the distance between the outermost views unchanged from the 5×5 views. In the second configuration (8×8 views (far)), the viewpoint intervals were kept unchanged from those for the 5×5 views, which made the distance between the outermost views 7/4 times larger. As shown in Fig. B.2, the performance with the 8×8 views (near) was almost equivalent to that with the 5×5 views, which indicates the generality of our method. Meanwhile, the performance with the 8×8 views (far) degraded to a certain degree, due to the larger angular range compared with the 5×5 views.

References

[1] Computer Graphics Laboratory, Stanford University. The (new) stanford light field archive, 2018. http://lightfield.stanford.edu. 1

[2] Heidelberg Collaboratory for Image Processing. Datasets and benchmarks for densely sampled 4D light fields. http://lightfieldgroup.iwr.uni-heidelberg.de/?page_id=713, 2016. 1

[3] Heidelberg Collaboratory for Image Processing. 4D light field dataset, 2018. http://hci-lightfield.iwr.uni-heidelberg.de/. 1

[4] MIT Media Lab's Camera Culture Group. Compressive light field camera. http://cameraculture.media.mit.edu/projects/compressive-light-field-camera/. 1

[5] Kohei Sakai, Keita Takahashi, Toshiaki Fujii, and Hajime Nagahara. Acquiring dynamic light fields through coded aperture camera. In *European Conference on Computer Vision*, pages 368–385, 2020. 2