Supplementary Material:
Motion parallax for 360° RGBD video

Ana Serrano[1]   Incheol Kim[1]   Zhili Chen[2]   Stephen DiVerdi[2]   Diego Gutierrez[1]

Aaron Hertzmann[2]   Belen Masia[1]

[1] Universidad de Zaragoza, I3A   [2] Adobe Research

This document offers additional information and details on the following topics:

- Depth improvement for motion parallax (Section 4 in the main paper)
- Layered video representation (Section 3 in the main paper)
- Results (Section 6 in the main paper)
- Evaluation (Section 5 in the main paper)

A   Depth improvement for motion parallax

We show more analyses of influence of the parameters in Equation 4 in the main paper, and the edge stopping function used for the edge guidance term (Equation 6 in the main paper).

A.1   Parameter Analysis

We show the influence of each term in Equation 4 in the main paper (Figure S.1). The first two rows show the results from different parameter sets of the data term (Equation 5). As the data parameter λ_{data} increases, the result gets similar to the input depth map that has artifacts. On the other hand, if λ_{data} is too small, our method overly corrects the input depth, resulting in wrong depth. The smoothness term (Equation 7) enforces smoothness of depth values, so the higher the parameter λ_{sm} is, the smoother the result becomes. However, if λ_{sm} is not enough large, our method fails to correct the regions in which wrong depth inputs should be smoothed out (See the region that is indicated by a white arrow in the case of $\lambda_{sm} = 5 \cdot 10^{-3}$). The edge guidance term takes edge-aware propagation into account. If λ_e is too small, bleeding/blurry artifacts cannot be removed well, and if it is too large, it overly corrects the input depth, resulting in wrong depth.

A.2   Edge-stopping Function

The edge guidance term (Equation 6 in the main paper) includes a spatially-varying weight $w_e(i, j)$ consisting of an edge-stopping function: Tukey's biweight. The edge-stopping function prevents propagation across edges, so the choice of it is crucial in our problem. The colorization method of Levin et al. [2004] uses an exponential function to give edge-awareness:

$$f_{\exp}(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{S.1}$$

where σ is a 3×3 local standard deviation. The anisotropic diffusion method of Perona and Malick [1990], they employ Lorentzian edge-stopping function:

$$f_{lorentzian}(x) = \frac{1}{1 + \frac{x^2}{2\sigma^2}}. \tag{S.2}$$

Black et al. [1998] suggest using Tukey's biweight as an edge-aware function:

$$f_{tukey}(x) = \begin{cases} \left(1 - \frac{x^2}{\sigma^2}\right)^2 & |x| \le \sigma \\ 0 & otherwise \end{cases}. \tag{S.3}$$

For all the functions in depth improvement, x is a local gradient value in 3×3 neighborhoods, which is denoted as $e(i) - e(j)$ where e is either an RGB or a depth pixel, i is an index of a pixel of interest, and j is i's local neighborhood.

Figure S.2 shows plots of the three functions with $\sigma^2 = 0.5$. All the plots show that around zero they have value close to one, and they rapidly drop in the other way around. In depth cleaning, x-axis is a gradient value of a guidance image, so we can
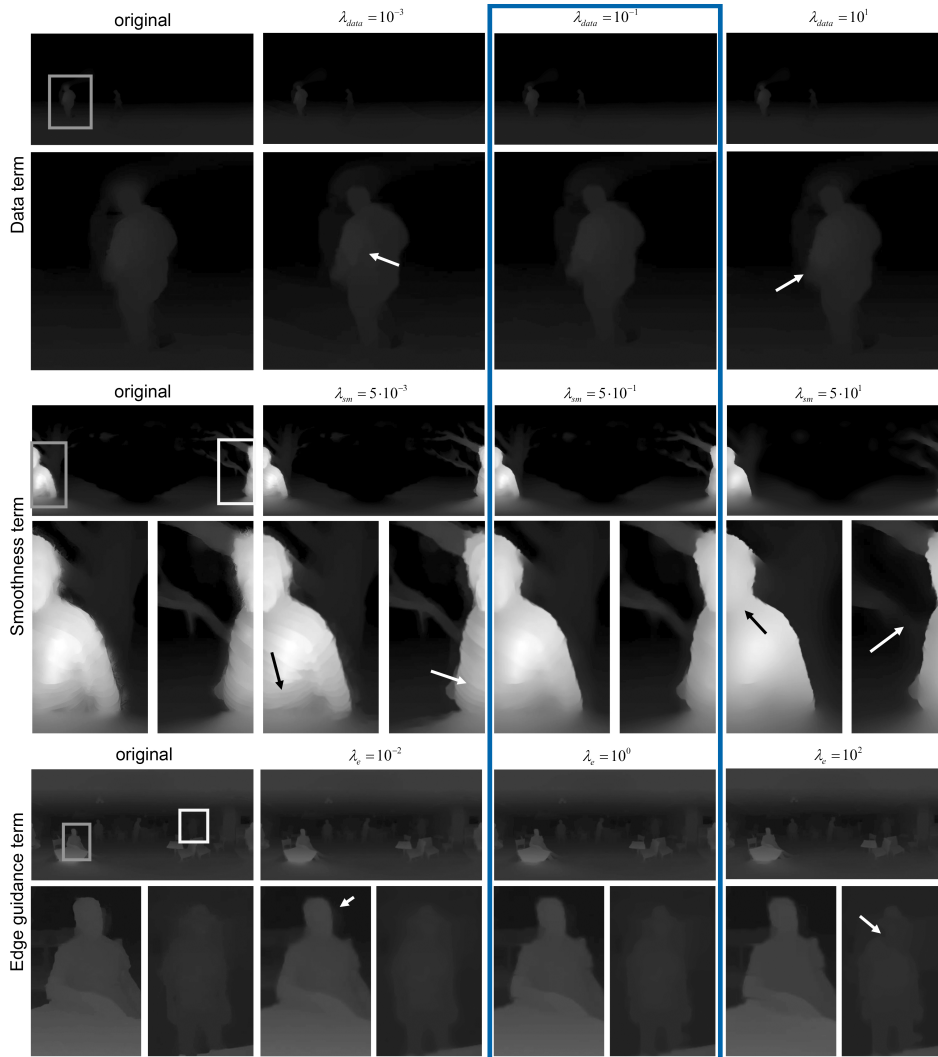
Figure S.1: *Results with different parameters. The third column is our parameter choice (marked with a blue square).*

see that if an edge-stopping function has a small value, it is likely that there is a prominent gradient (edge), and vice versa. Lorentzian function shows the most generous behavior in determining edges since it has the thickest tails. The exponential function plummets as it goes away from zero, but Tukey's biweight used in our method shows the strictest determinant of gradients. In other words, if there is even a small gradient, Tukey's biweight blocks propagation during optimization while Lorentzian allowing. The double arrows in Figure S.2 indicate the intervals that the functions allow propagation. Figure S.3 shows the results from different edge-stopping functions. Lorentzian function shows the smoothest result since its gradient determinant is generous. The exponential function gives better results, but around non-prominent edges, it shows blurry depth boundaries. Tukey's biweight produces the sharpest outputs since its gradient threshold is the strictest.
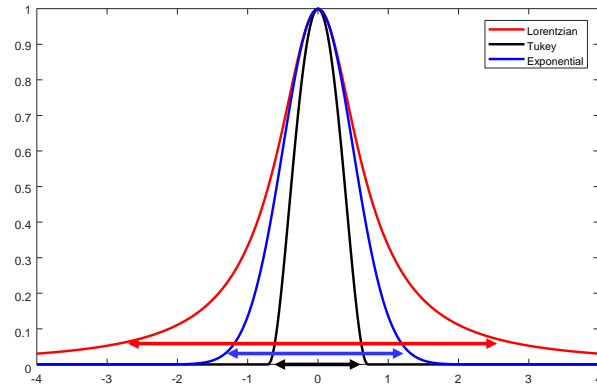
Figure S.2: *Edge stopping functions. For our results we choose the Tukey stopping function.*
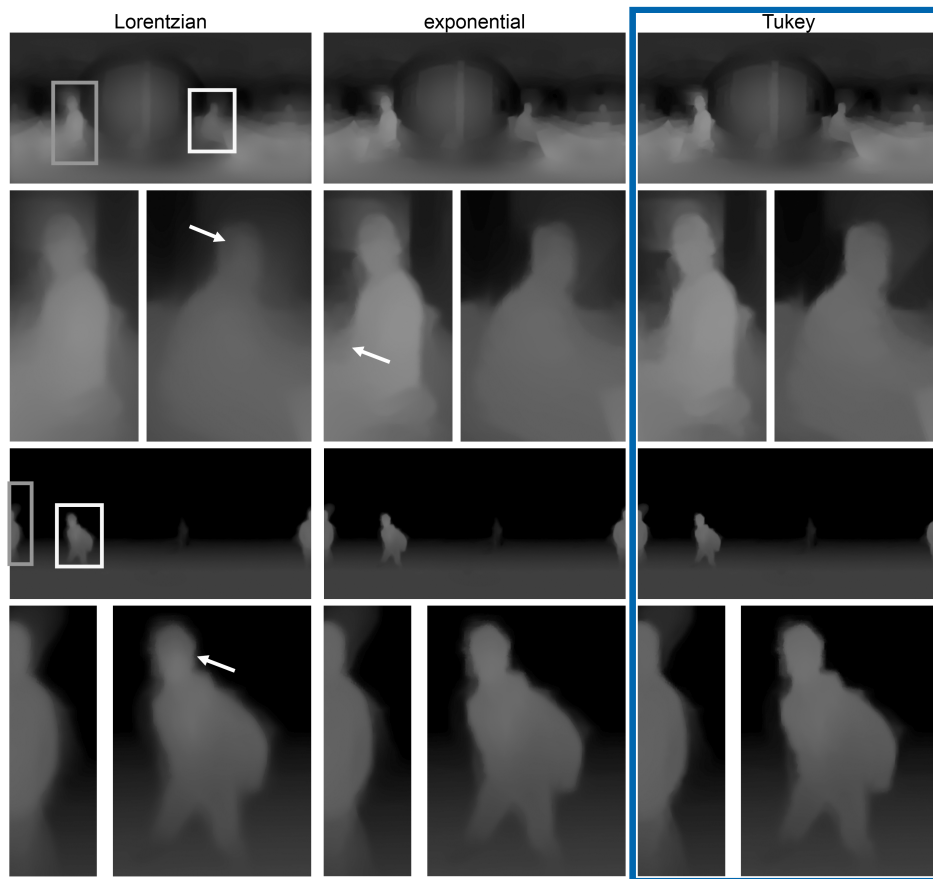

Figure S.3: *Results from different edge-stopping functions. For our results we choose the Tukey stopping function (marked with a blue square).*

B   Layered video representation

B.1   Computation of opacity map α: Intermediate steps

We show a more detailed pipeline of our α processing explained in Section 3 in the main paper. See Figure S.4 for the description of the pipeline.
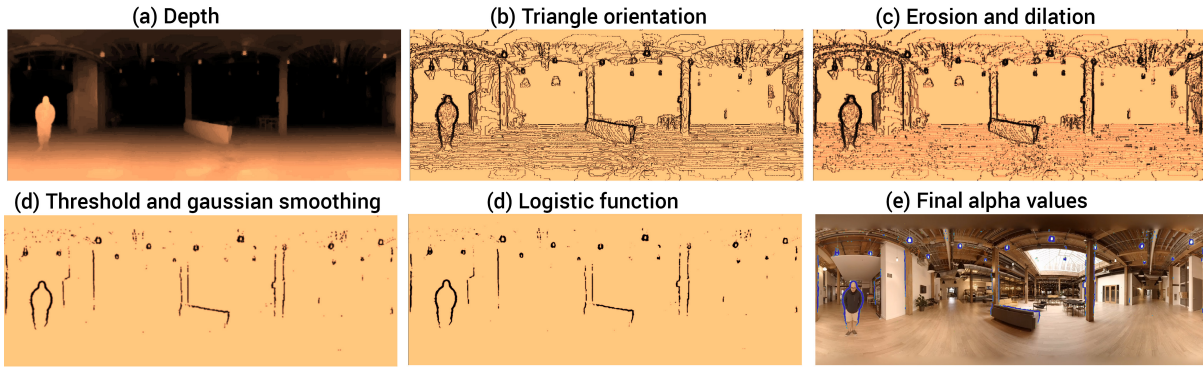
Figure S.4: *(a) Depth map of the foreground layer for a given frame of the video. (b) Raw fragment orientation map O^F. (c) Intermediate orientation map processed with erosion and dilation operators ($O^F \bullet K$). (d) Processed fragment orientation map $\hat{\alpha}^F$. The raw orientation values ($O^F_{i,j}$) are too noisy to provide smooth transparency values, as they include information not only on disocclusions, but also on foreshortening and quantization errors. We process this raw values as described in the main paper, and obtain a clean transparency map ($\hat{\alpha}^F$) that smoothly matches the RGB image. (e) RGB image of which the processed fragment orientation map $\hat{\alpha}^F$ is overlaid on top.*

B.2 Computation of maximum parallax ρ

In this section we include additional details on how to compute the maximum parallax angle ρ, this dictates the number of pixels to inpaint in the *inpainted* background layer. The geometrical representation for the following derivation is given in Figure S.5. We provide the derivation for the 2D case for simplicity, since the extension for 3D space is straight-forward. Given
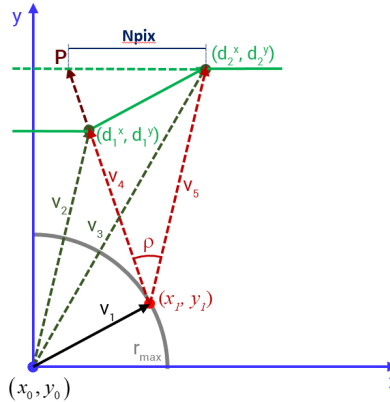
Figure S.5: *Illustration that accompanies the computation of maximum parallax ρ (see main supplementary text).*

two consecutive mesh vertices d_1 and d_2, the maximum parallax ρ depends on the depth difference between such vertices, and the maximum head movement allowed (r_{max}). The vectors v_2 and v_3 are given by the position of the two vertices, (d_1^x, d_1^y) and (d_2^x, d_2^y). The vector v_1 is determined by setting a maximum head displacement circle around the center of projection (x_0, y_0) with radius r_{max}. The angle of maximum parallax ρ can be expressed as:

$$
\begin{aligned}
\cos(\rho) &= \frac{v_4 \cdot v_5}{\|v_4\|\|v_5\|}, \\
with \\
v_5 &= v_3 - v_1, \\
v_4 &= v_2 - v_1
\end{aligned}
\tag{S.4}
$$

We aim to compute the maximum ρ, so we can formulate our problem as a maximization problem in the form:

$$
\underset{x_1, y_1 \leq r}{\arg\max} \, \rho = \arccos \frac{v_4 \cdot v_5}{\|v_4\|\|v_5\|}
\tag{S.5}
$$

If we substitute v_4 and v_5 by their expressions as defined in Equation S.4:

$$\arg\max_{x_1, y_1 \leq r} \arccos \frac{(d_x^2 - x_1)(d_x^1 - x_1) + (d_y^2 - y_1)(d_y^1 - y_1)}{\sqrt{(d_x^2 - x_1)^2 + (d_y^2 - y_1)^2}\sqrt{(d_x^1 - x_1)^2 + (d_y^1 - y_1)^2}} \tag{S.6}$$

Finally, we can assume that maximum angle of parallax ρ will lie in the outer boundary, such that $x_1^2 + y_1^2 = r_{max}^2$:

$$\arg\max_{|x_1| \leq r_{max}} \arccos \frac{(d_x^2 - x_1)(d_x^1 - x_1) + (d_y^2 - \sqrt{r^2 - x_1^2})(d_y^1 - \sqrt{r^2 - x_1^2})}{\sqrt{(d_x^2 - x_1)^2 + (d_y^2 - \sqrt{r^2 - x_1^2})^2}\sqrt{(d_x^1 - x_1)^2 + (d_y^1 - \sqrt{r^2 - x_1^2})^2}} \tag{S.7}$$

where all the terms but x_1 are known, and we seek to find x_1 such that $|x_1| \leq r_{max}$. By knowing x_1 and y_1 we can go back to Equation S.5 and obtain the maximum angle of parallax ρ. By knowing ρ, we can compute P (assuming a planar surface to inpaint, i.e., for this example this will imply that P lies at the same distance than the further mesh vertex d_2 in the y axis) to calculate the number of pixels N_{pix} to inpaint in the *inpainted* layer.

C   Results

We show additional examples of results as described in Section 6 of the main paper. Six results are shown in Figure S.6, depicting a variety of scenes. For each result, we show a view from the center of projection, and a displaced view leading to disocclusions. In each of the scenes we highlight regions illustrating the added parallax. We also include results from monocular videos from different sources with estimated depth in Figure S.7.
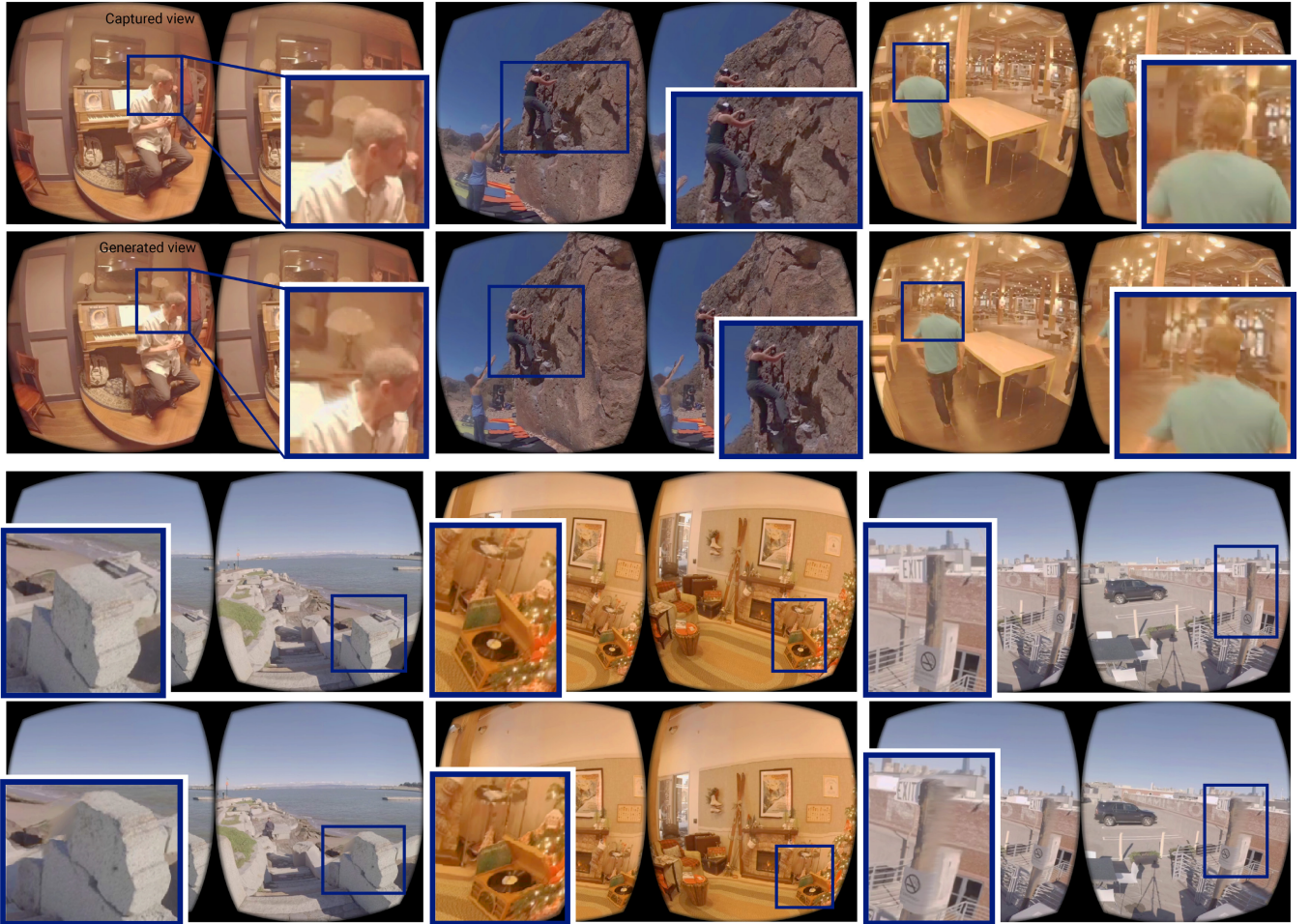
Figure S.6: *Six examples of novel views generated with our method. For each example we show the original view (top), and the corresponding displaced view (bottom). We also include close-ups of regions where the added parallax is clearly visible.*
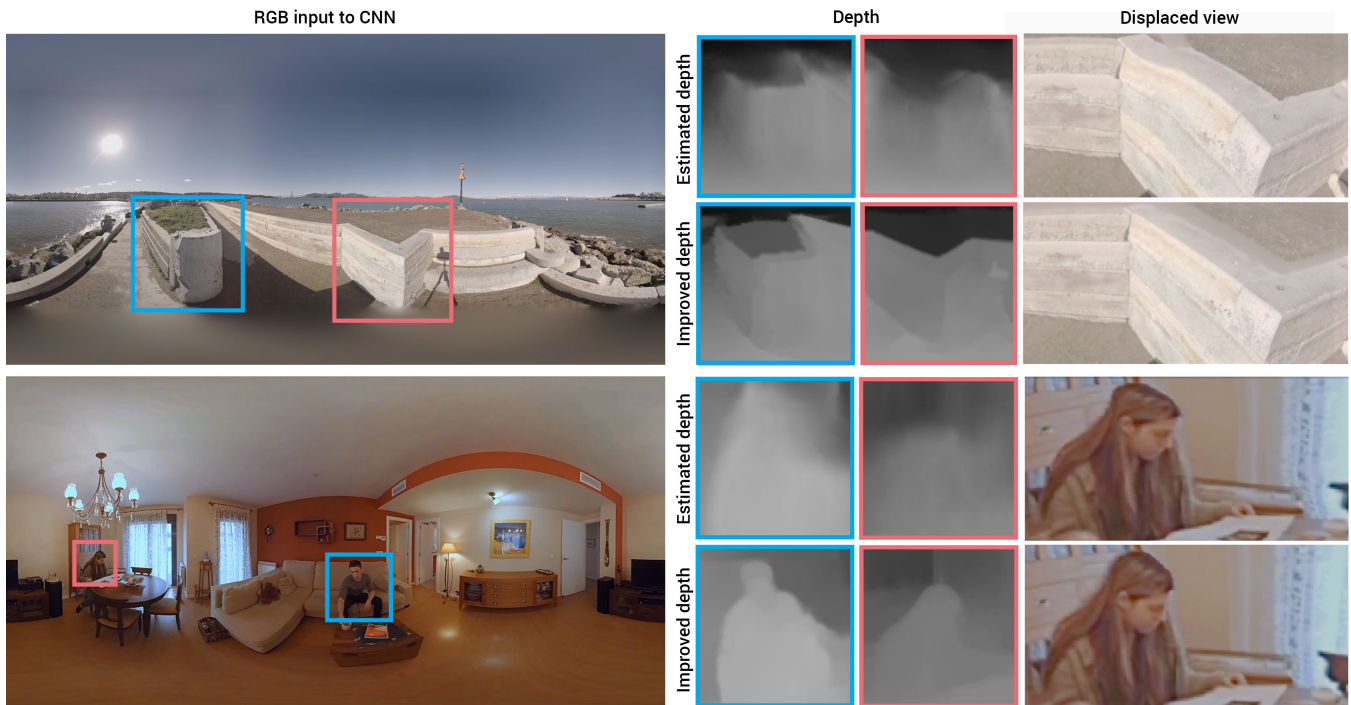
Figure S.7: *Results using monocular video without depth as input. For each row, we show a representative frame (left), details of the initial estimated depth [Godard et al. 2017] and our improved depth (middle), as well as the corresponding results when generating a novel view (right). Our method yields minimal distortions even in the presence of such suboptimal input. The first row depicts a scene captured for this paper with a GoPro Odyssey, while the second row is taken from a previously existing short clip (dataset from [Serrano et al. 2017]); in both cases we use only a 360° RGB monocular view as input and drop the remaining data.*

D Evaluation

We show additional plots from the experiments explained in Section 5 of the main paper, and the questionnaires used for user study.
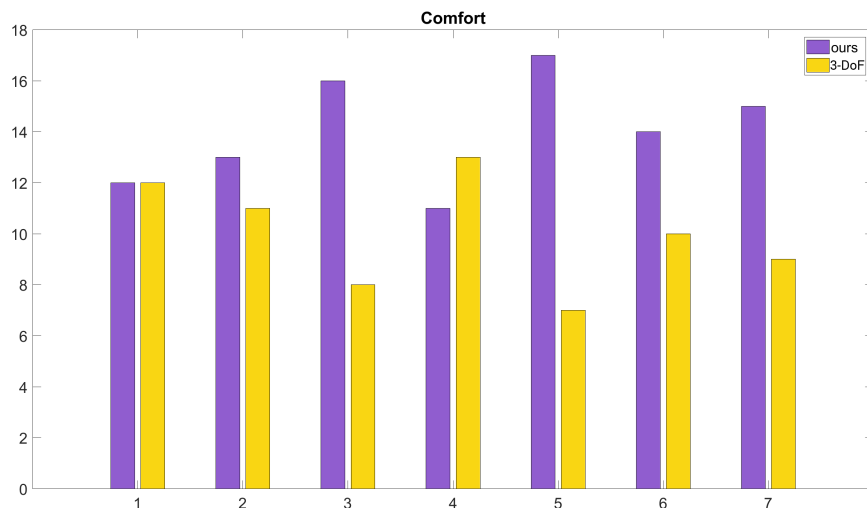
D.1 Experiment #1

We conduct a user study without noticing participants about the difference between two visualizations modes: with/without head-motion parallax (Experiment #1 in the main paper). Figure S.8 shows representative frames of the dataset used for this user study. We show the vote counts of artifacts (Figure S.9), comfort (Figure S.10), immersion (Figure S.11), preference (Figure S.12), and realism (Figure S.13).
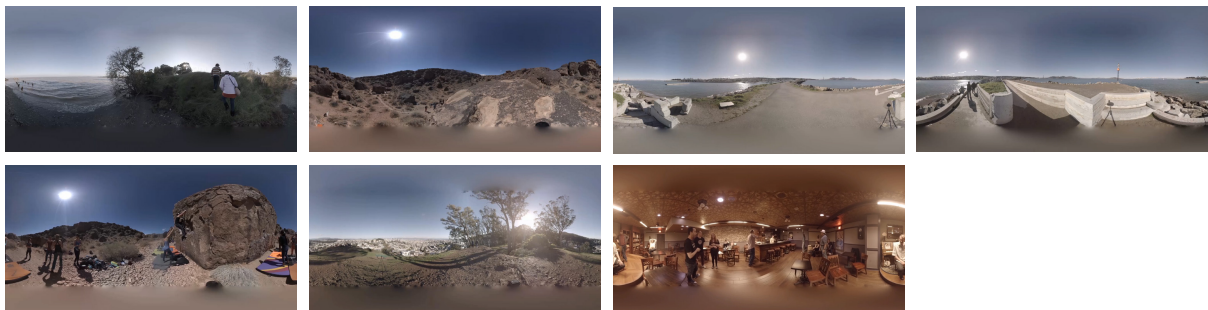
Figure S.8: *Representative frames of Experiment #1.*
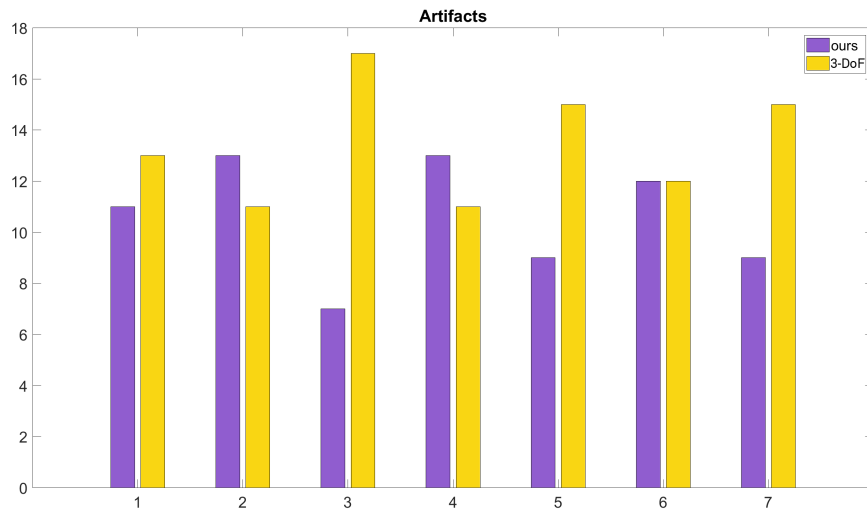
Figure S.9: *Votes of artifacts in Experiment #1 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The lower a vote count is, the less the artifacts are seen by a participant.*
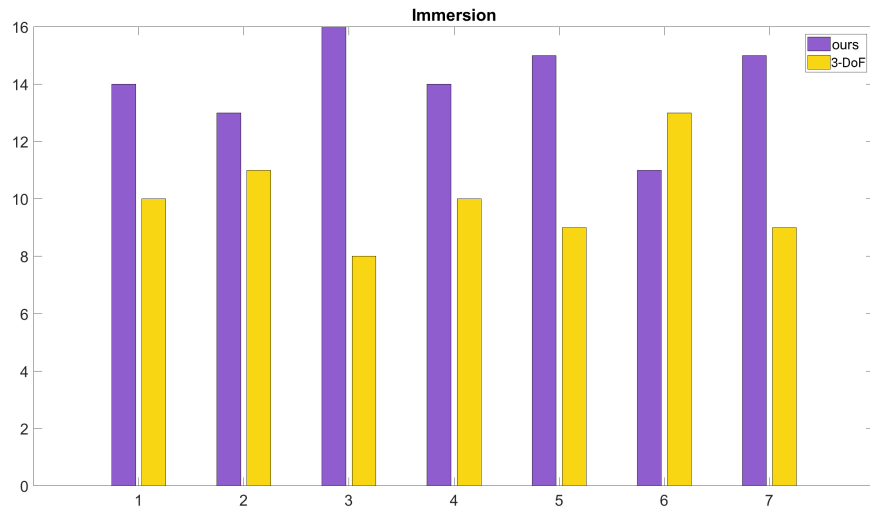

Figure S.10: *Votes of comfort in Experiment #1 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The higher a vote count is, the more a participant prefers.*
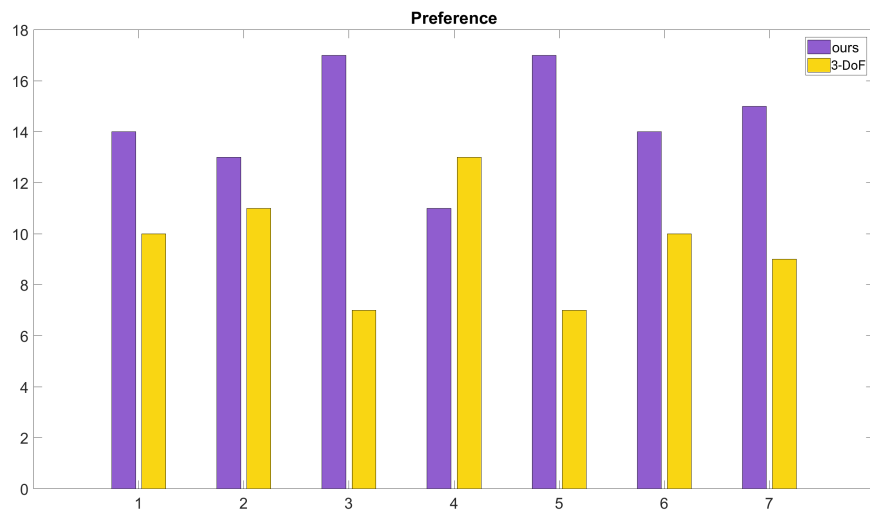
D.2 Experiment #2

We conduct a user study to check sickness in VR depending on whether motion parallax is supported or not (Experiment #2 in the main paper). Figure S.14 shows representative frames of the dataset used for this user study.

D.3 Experiment #3

We conduct a user study with noticing participants about the difference between two visualizations modes: with/without head-motion parallax (Experiment #3 in the main paper). Figure S.15 shows representative frames of the dataset used for this user study. We show the vote counts of artifacts (Figure S.16), comfort (Figure S.17), immersion (Figure S.18), preference (Figure S.19), and realism (Figure S.20).
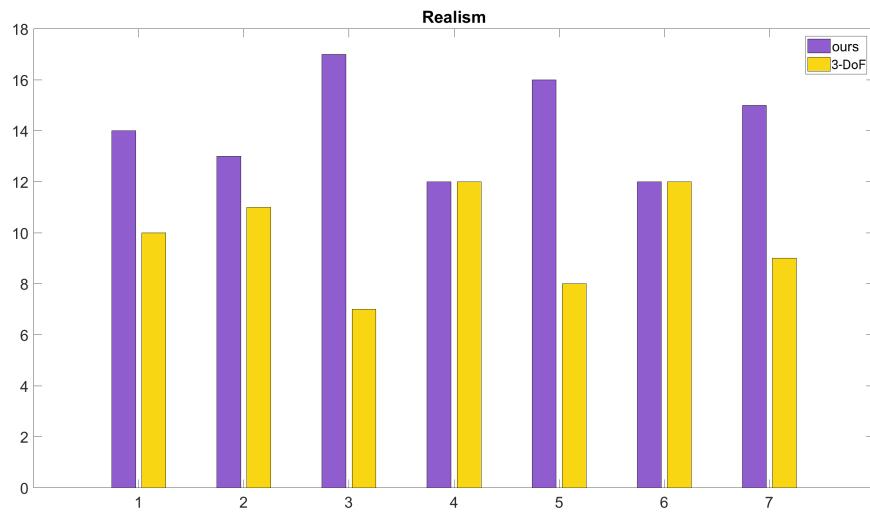
Figure S.11: *Votes of immersion in Experiment #1 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The higher a vote count is, the more a participant prefers.*

Figure S.12: *Votes of preference in Experiment #1 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The higher a vote count is, the more a participant prefers.*

Figure S.13: *Votes of realism in Experiment #1 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The higher a vote count is, the more a participant prefers.*

Figure S.14: *Representative frames of Experiment #2.*

Figure S.15: *Representative frames of Experiment #3 .*

Figure S.16: *Votes of artifacts in Experiment #3 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The lower a vote count is, the less the artifacts are seen by a participant.*

Figure S.17: *Votes of comfort in Experiment #3 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The higher a vote count is, the more a participant prefers.*
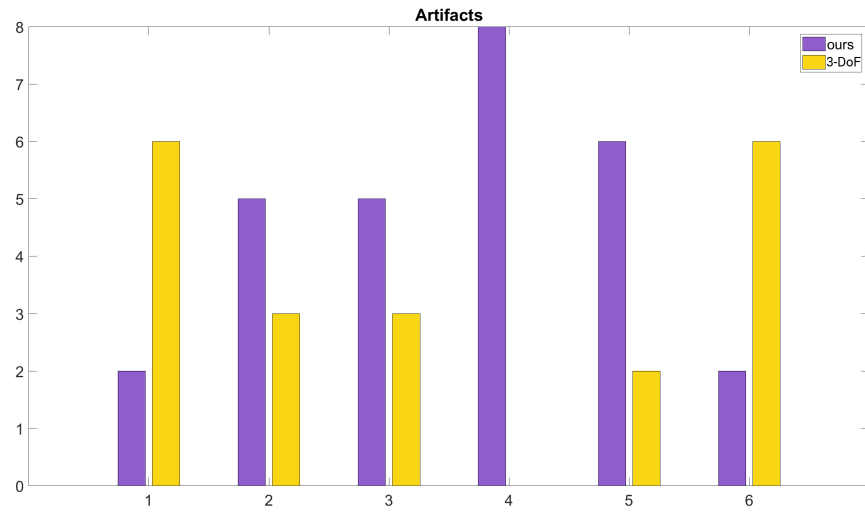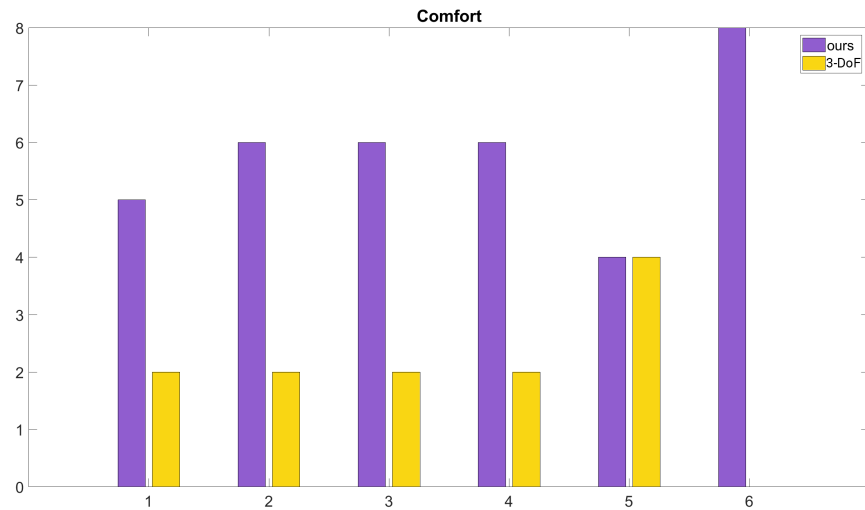
Figure S.18: *Votes of immersion in Experiment #3 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The higher a vote count is, the more a participant prefers.*
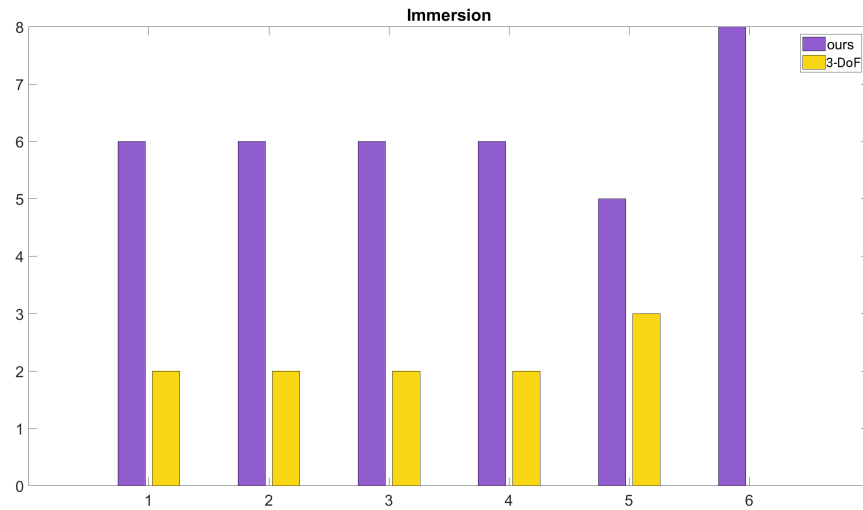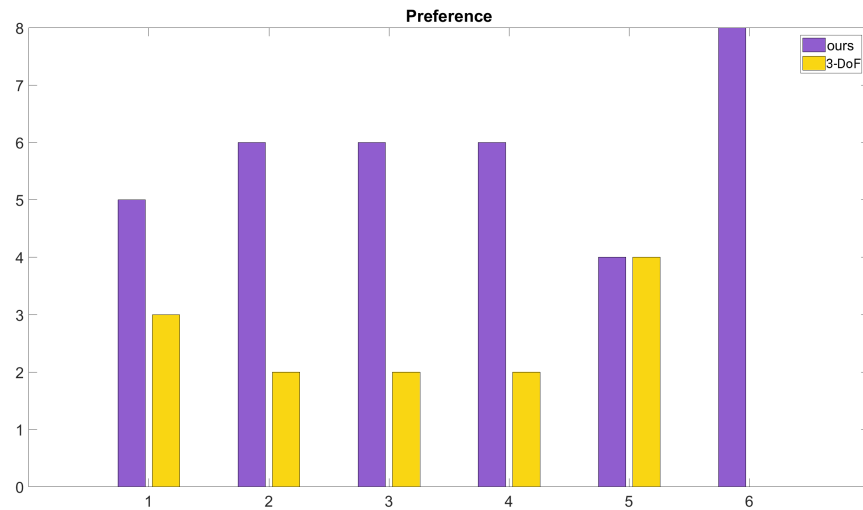

Figure S.19: *Votes of preference in Experiment #3 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The higher a vote count is, the more a participant prefers.*
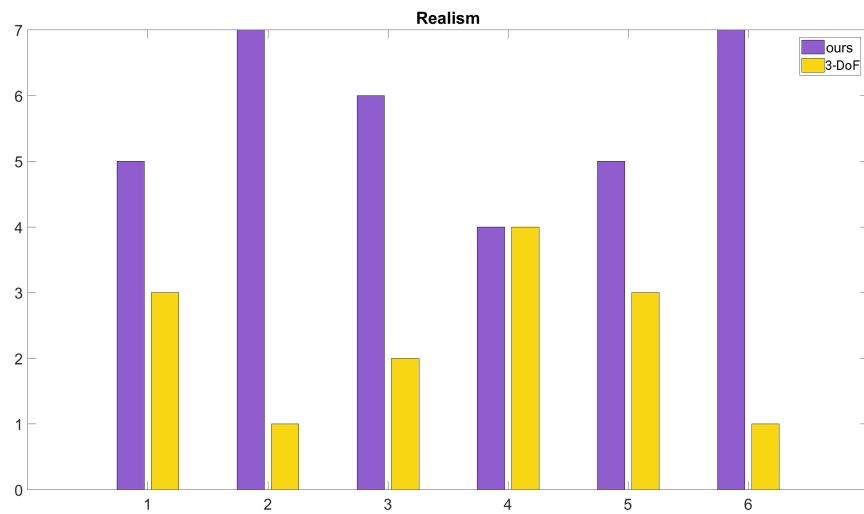
Figure S.20: *Votes of realism in Experiment #3 for our method with motion parallax and the method without motion parallax. The purple bars indicate the vote counts of our method (6-DoF), and the yellow bars show those of 3-DoF viewer. The higher a vote count is, the more a participant prefers.*

D.4 Questionnaires

We include the questionnaires used for our user study. The first questionnaire is to ask participants if they are familiar with a VR headset, prior to the experiments. The second questionnaire is used to ask preference of one viewing mode over the other in terms of realism, comfort, immersion, presence of visual artifacts, and global preference. To assess if the subjects experienced sickness, dizziness, and/or vertigo, and whether they had experienced discomfort, and if so, to measure how much they felt those during viewing, we use the sickness questionnaire.

EXPERIMENT ID _____

Age _____
Gender _____

Have you used an HMD before?   YES / NO
If yes:

      Was it PC-based or smartphone-based?
      How many times have you used it? _____
      When was the last time you used it?_____
      Do you use it regularly (more than once a month)?   YES / NO
      Have you ever experienced eyestrain, dizziness, headaches, or nausea in VR?   YES / NO

Do you play videogames regularly?   YES/NO

If yes, how often (days/week)?_____

Other information?_____

Experiment #1 and Experiment #3

EXPERIMENT ID   _____
RENDERING MODE   _____
VIDEO _____

1. With which of the two methods did you feel more immersed in the scene?

       First method ☐   Second method ☐

Comments:………
………………………………………………………………………………………………………..

2. Which of the two methods had more visual artifacts?

       First method ☐   Second method ☐

Comments:………
………………………………………………………………………………………………………..

3. Which of the two methods offers a more realistic experience?

       First method ☐   Second method ☐

Comments:………
………

4. With which of the two methods did you feel more comfortable watching the scene?

       First method ☐   Second method ☐

Comments:………
………

5. Globally, which of the two methods do you prefer for visualizing the scene?

       First method ☐   Second method ☐

Comments:………
………

6. [This is a simple question regarding the scene content, it differs for each scene]

………

SICKNESS QUESTIONNAIRE (Experiment #2)

EXPERIMENT ID   _____

RENDERING MODE _____

General discomfort
  ○   ○   ○   ○
None Slight Moderate Severe

Fatigue
  ○   ○   ○   ○
None Slight Moderate Severe

Eyestrain
  ○   ○   ○   ○
None Slight Moderate Severe

Difficulty focusing
  ○   ○   ○   ○
None Slight Moderate Severe

Headache
  ○   ○   ○   ○
None Slight Moderate Severe

Fullness of head
  ○   ○   ○   ○
None Slight Moderate Severe

Blurred vision
  ○   ○   ○   ○
None Slight Moderate Severe

Dizzy (eyes closed)
  ○   ○   ○   ○
None Slight Moderate Severe

Vertigo
  ○   ○   ○   ○
None Slight Moderate Severe

Did you feel totally comfortable while watching the videos?

_____

Did you experience at any moment while watching the videos any symptom of dizziness, vertigo, disorientation, or nausea? If yes, describe the video and situation where this happened.

_____

References

BLACK, M. J., SAPIRO, G., MARIMONT, D. H., AND HEEGER, D. 1998. Robust anisotropic diffusion. *IEEE Trans. Image Processing 7*, 3, 421–432.

GODARD, C., MAC AODHA, O., AND BROSTOW, G. J. 2017. Unsupervised monocular depth estimation with left-right consistency. In *CVPR*, IEEE Computer Society, 6602–6611.

LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2004. Colorization using optimization. *ACM Trans. Graph. 23*, 3, 689–694.

PERONA, P., AND MALIK, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell. 12*, 7, 629–639.

SERRANO, A., SITZMANN, V., RUIZ-BORAU, J., WETZSTEIN, G., GUTIERREZ, D., AND MASIA, B. 2017. Movie editing and cognitive event segmentation in virtual reality video. *ACM Transactions on Graphics (SIGGRAPH 2017) 36*, 4.