Memory Efficient Stereoscopy from Light Fields
Supplementary Material

Changil Kim[1,2] Ulrich Müller[1,2] Henning Zimmer[1] Yael Pritch[1]
Alexander Sorkine-Hornung[1] Markus Gross[1,2]

[1]Disney Research Zurich [2]ETH Zurich

S1. Introduction

We present the complete set of results including those omitted in the paper, and further elaborate on the relation between our method and the dense disparity estimation problem.

S2. More Results and Comparisons

Figures S1–S4 show the complete set of input and output images of the four tasks we experimented using four different datasets at various resolutions. Since the Elephant dataset was available only at 1k resolution without depth information, we only conducted the *user scribbles* and the *constant disparity* tasks at 1k resolution for this dataset. Please also refer to our paper for more details of the experiments.

Table S1 lists the root-mean-squared errors (RMSE) of our results against the ground truth for the *constant disparity* and *linear scaling* tasks, where the new view is computed with respect to a constant disparity and a linear scaling of the scene depth, respectively. We measured the error for two different resolutions, 1280×853 (1k) and 1920×1280 (2k). Refer to Figures S3 and S4 for the corresponding results.

Target disparity	RMSE from the ground truth			
	Elephant	Bikes	Couch	Mansion
1k resolution				
Constant	0.0499	0.0396	0.0235	0.0704
Linear	N/A*	0.0315	0.0072	0.0774
2k resolution				
Constant	N/A*	0.0392	0.0252	0.0735
Linear	N/A*	0.0288	0.0065	0.0693

Table S1. *Errors of the computed views against the ground truth.* (*Dataset/depth not available)

Table S2 summarizes the memory consumption of our method for a set of several different light field resolutions. We also compare it against the current state of the art [1] which uses a discrete graph-cut formulation. All numbers were measured using our implementation. We were not able to measure the memory footprint of the discrete formulation on 2k resolution datasets because the test system became unresponsive due to the excessive page swapping.

Resolution (w×h×#images)	#pixels (Mpix)	Memory use in Mb		Ratio
		Ours	Discrete	
640×427×30	8.2	186.9	1,761.0	9.4%
640×427×50	13.7	310.0	2,983.4	9.2%
1k resolution				
1280×853×30	32.8	664.3	7,047.3	9.4%
1280×853×50	54.6	1,101.7	11,941.3	9.2%
2k resolution				
1920×1280×30	73.7	1,495.3	N/A*	–
1920×1280×50	122.9	2,479.7	N/A*	–

Table S2. *Memory footprint of our method* in comparison to the state-of-the-art discrete formulation [1]. (*Test failed)

Table S3 shows the running time of our method for all four tasks and two different resolutions. We used the same,

Target disparity	Computation time in seconds			
	Elephant	Bikes	Couch	Mansion
1k resolution				
Constant	§946	†513	660	614
Linear	N/A*	†518	667	623
Scribbles	636	†514	661	616
Remapping	N/A*	†516	669	628
2k resolution				
Constant	N/A*	838	863	803
Linear	N/A*	842	860	814
Scribbles	N/A*	840	858	799
Remapping	N/A*	837	860	805

Table S3. *Running time of our method.* We used 50 1k images or 30 2k images for the measurements. (*Dataset/depth not available; §70 images used; †40 images used) fixed set of parameters for all experiments, including the number of primal-dual iterations.

S3. Relation to the Dense Stereo Problem

The labeling results of the *constant disparity* task in Figure S3 resemble the actual scene depth. When a constant disparity g is used as the target disparity, i.e., $G(\mathbf{x}) = g$, $\forall \mathbf{x} \in \Omega$, the mapping (1) between the reference image $I_{\hat{s}}$ and the target image $I_{\hat{s}}^*$,

$$I_{\hat{s}}^*(u + G(u,v), v) = I_{\hat{s}}(u,v)\,, \tag{S1}$$

becomes bijective, and $M(\mathbf{x}) = 1$ everywhere except for the g-pixel-wide vertical strip at the left image border. The data term (3), copied in (S3) below, can thus be rewritten as

$$\|L(u,v,l(u,v)) - L(u-g,v,\hat{s})\|_1\,, \tag{S2}$$

for a pixel $\mathbf{x} = (u,v) \in \Omega : u > g$.

Minimizing this energy, together with the smoothness term, for l finds for each pixel (u,v) the matching two features at $(u,v,l(u,v))$ and $(u-g,v,\hat{s})$, from which the disparity is computed as $g/(l(u,v)-\hat{s})$ using a simple triangulation. Since both g and $\hat{s}$ are constant, the labeling l looks like the disparity map. In fact, the data term (3) in our paper,

$$M(u,v)\,\|L(u,v,l(u,v)) - I_{\hat{s}}^*(u,v)\|_1\,, \tag{S3}$$

implicitly implements dense disparity estimation, which can be shown clearer using the *linear scaling* task below.

With the (scaled) actual disparity as the target disparity one obtains a flat labeling as shown in Figure S4. Let us assume that the target disparity G gives us an injective mapping from the reference image $I_{\hat{s}}$ to the target image $I_{\hat{s}}^*$ in (S1). Plugging this mapping (S1) into the data term (S3) then yields

$$M(u,v)\,\|\,L(u,v,l(u,v)) - L(u-G(u,v),v,\hat{s})\,\|_1\,. \tag{S4}$$

In our original problem, we fix the target disparity G and seek the image index l for each pixel $(u,v) \in \Omega$. If, instead, we fix the labeling l to be a constant s' over Ω (i.e., flat labeling), and optimize the functional for G over all pixels $(u,v) \in \Omega$, the result will be the disparity map defined between the two images at the reference image $\hat{s}$ and the fixed other view s'. In this case the smoothness should accordingly be redefined in terms of G, instead of l.

References

[1] C. Kim, A. Hornung, S. Heinzle, W. Matusik, and M. H. Gross. Multi-perspective stereoscopy from light fields. *ACM Trans. Graph.*, 30(6):190, 2011. 1

(Figures follow on next page.)

(a) User scribbles (b) Propagated disparity (c) Reference image (d) Computed new view (e) Resulting stereo

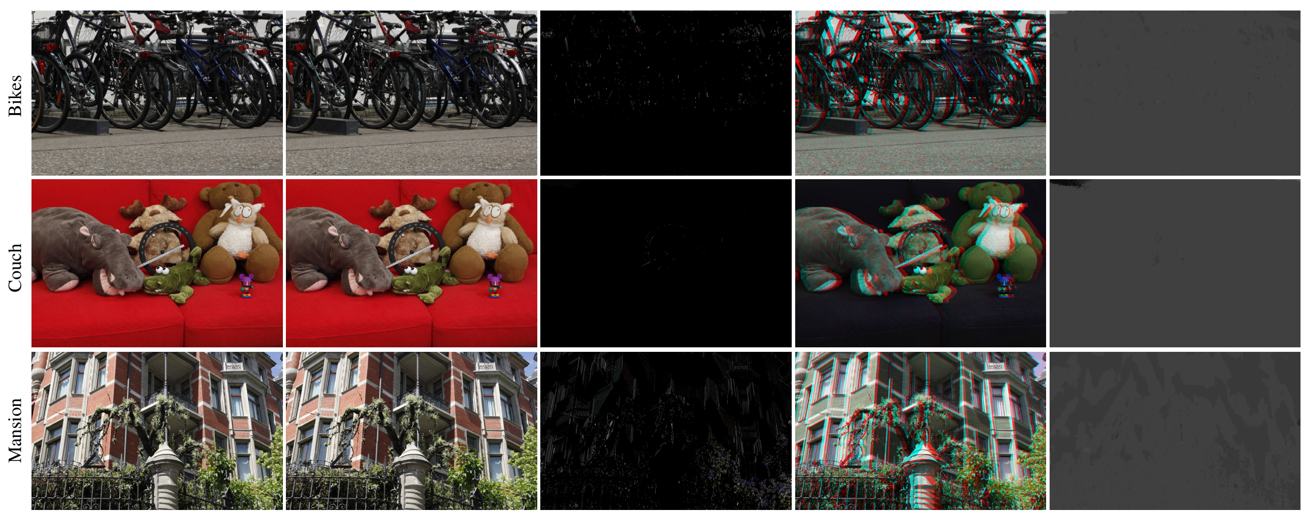
Figure S1. *Disparity modification using user scribbles.* This task demonstrates a possible use case, where sparse brush strokes are drawn by the user (a) and then propagated to form a dense target disparity map (b) from which the resulting stereo is generated. (c) and (d) show the reference view and the computed new view, respectively. (e) shows the resulting anaglyph stereo image. Note that the scribbles are not necessarily physically meaningful and are rather intended to test the flexibility and robustness of our method.

(a) Original disparity (b) Remapped disparity (c) Reference image (d) Computed new view (e) Resulting stereo

Figure S2. *Nonlinear disparity remapping.* The actual scene depth of the reference view (a) is nonlinearly remapped to create the target disparity map (b). For the Bikes dataset, the excessive disparity on the ground was compressed for a more comfortable stereoscopic viewing experience. For the Couch and Mansion datasets, the gradient of the disparity is modified such that large disparity gradients are removed, to better distribute the disparity budget and to obtain more local details. (c–e) show the reference image, the computed new view, and the resulting anaglyph stereo image, respectively.

(a) Reference image (b) Computed new view (c) Error (d) Anaglyph stereo (e) Labeling

Figure S3. *Constant disparity.* (a) and (b) show the reference image and the computed new view given a fixed value of 20 pixels as the target disparity. (c) shows the error of the computed image against the ground truth, for which we use the reference image translated by 20 pixels. The darker the pixel in the error image, the smaller error. See Table S1 for corresponding RMSE measures. (d) shows the anaglyph stereo image, while (e) shows the resulting labeling. The resulting stereo should ideally look flat, but floating on the screen. The labeling images look like depth maps of the scenes. In fact, the problem we address and the dense disparity estimation problem are closely related; see Section S3 of this document.



(a) Reference image (b) Computed new view (c) Error (d) Anaglyph stereo (e) Labeling

Figure S4. *Linear disparity scaling.* (a) and (b) show the reference image and the computed new view, for which the depth at the reference view was linearly scaled by a factor of 10 and used as the target disparity map. (c) shows the error of the computed image against the ground truth, i.e., the 10th next image to the reference in the input light field. The darker the pixel in the error image, the smaller error. See Table S1 for corresponding RMSE measures. (d) shows the anaglyph stereo image, and (e) shows the resulting labeling. The labeling should ideally look flat in this task.