Single-shot Monocular RGB-D Imaging using Uneven Double Refraction (Supplemental Document)

Andreas Meuleman[1*] Seung-Hwan Baek[1*†] Felix Heide[2] Min H. Kim[1]

[1]KAIST [2]Princeton University

This document includes notation table, calibration details, proof for the rectification method, and detailed evaluation of our method.

1. Pseudo Code and Table

Algorithm 1 describes our image reconstruction process. Table 1 provides the symbols and notation used in the paper.

Algorithm 1 Our image restoration algorithm.

function ImageRestore($\tilde{I}_c, z$)
  $\hat{I}_o^{z,(0)} = \tilde{I}_c$
  $\hat{\Delta}^{z,(0)} = \tau A\left(\hat{I}_o^{z,(0)}, \tilde{r}_{o\to e}(z)\right)$
  for $n = 0$ to $N-1$
    $\hat{I}_o^{z,(n+1)} = \hat{I}_o^{z,(n)} - \hat{\Delta}^{z,(n)}$
    $\hat{\Delta}^{z,(n+1)} = -\tau^{2^{(n+1)}} \cdot A\left(\hat{I}_o^{z,(n+1)}, 2^{(n+1)} \cdot \tilde{r}_{o\to e}(z)\right)$
  end for
  return $\hat{I}_o^{z,(N)}$
end function

Symbol	Description
f	Focal length
z	Depth candidate
Z	Estimated depth
P_s	Scene point
P_o	Pixel hit by an ordinary ray
P_e	Pixel hit by an extraordinary ray
P_d	Pixel hit by a direct ray
P	A pixel within the image domain
T	Rectification transformation function
$r_{\overline{od}}$	Disparity from P_o to P_d
$r_{\overline{de}}$	Disparity from P_d to P_e
$r_{\overline{oe}}$	Disparity from P_o to P_e
$\hat{r}_{\overline{oe}}$	Disparity from P_o to P_e with depth invariance
$\tilde{r}_{\overline{oe}}$	Disparity from P_o to P_e with depth and spatial invariance
$b_{\overline{oe}}$	Birefractive baseline
$\hat{b}_{\overline{oe}}$	Baseline with depth invariance
$\hat{b}^{\text{avg}}_{\overline{oe}}$	Average of $\hat{b}_{\overline{oe}}$ along the horizontal axis
$\tilde{b}_{\overline{oe}}$	Baseline with depth and spatial invariance
I_c	Captured superimposed image
$\tilde{I}_c$	Rectified captured image
$\tilde{I}_o$	Rectified image component from ordinary rays
$\tilde{I}_e$	Rectified image component from extraordinary rays
$\hat{I}_o^{z,(n)}$	Restored image of $\tilde{I}_o$ at n^{th} iteration for a given depth candidate z
$\Delta^{z,(n)}$	Residual image we intend to eliminate from $\hat{I}_o^{z,n}$
$\hat{\Delta}^{z,(n)}$	Approximated residual image
$\hat{I}_o^z$	Restored image result of $\tilde{I}_o$ for a given z (slice of the restoration volume)
$\hat{I}_o^Z$	Restored image of $\tilde{I}_o$ with our estimated depth Z
τ	Intensity proportion between the e-ray and o-ray
C	Stereo matching cost

Table 1. Symbols and notations used in the paper.

2. Proof for Depth-invariant Baseline

We describe our observation and proof for the approximation of the original birefractive baseline $b_{\overline{oe}}$ with the depth-independent baseline $\hat{b}_{\overline{oe}}$:

$$\frac{f}{z} b_{\overline{oe}}(P_o, z) \approx \frac{f}{z} \hat{b}_{\overline{oe}}(P_o).$$

Observation We first describe the original birefractive stereo model using the baseline $b_{\overline{oe}}$ as follows:

$$\frac{f}{z} b_{\overline{oe}}(P_o, z) = \frac{f}{z} b_{\overline{od}}(P_o) + \frac{f}{z} b_{\overline{de}}\left(P_o + r_{\overline{od}}(P_o, z)\right). \quad (1)$$

We analyze the depth-dependent term in the baseline of the original model: $b_{\overline{de}}(P_o + r_{\overline{od}}(P_o, z))$ in Equation (1) w.r.t. our hardware configuration. Given hardware parameters of the focal length, calcite thickness, o-/e-ray refractive indices of calcite, we made two observations on disparity and baseline. First, the maximum magnitude of $b_{\overline{od}}$ is bounded:

$$\left\| b_{\overline{od}}(P_o) \right\|_2 < \alpha = 485, \forall P_o.$$

Second, the difference of $b_{\overline{de}}$ of two closest points, $(P_o + Q_o)$ and P_o, is bounded:

$$\left\| b_{\overline{de}}(P_o + Q_o) - b_{\overline{de}}(P_o) \right\|_2 < \beta = 0.28, \forall P_o, Q_o \quad \text{s.t. } \|Q_o\|_2 \leq 1.$$

This indicates the local consistency of $b_{\overline{de}}$ that can be generalized to any point Q_o as:

$$\left\| b_{\overline{de}}(P_o + Q_o) - b_{\overline{de}}(P_o) \right\|_2 < \beta \|Q_o\|_2.$$

Proof We can apply our observations on both $b_{\overline{od}}$ and $b_{\overline{de}}$ to detach the depth-dependency of the birefractive base-

∗: Equal contribution. †: Now at Princeton University

line $b_{\overline{oe}}$. Specifically, we extend our observation on the local consistency of $b_{\overline{de}}$ by setting Q_o as $r_{\overline{od}}(P_o, z)$ with the scaler of $\frac{f}{z}$ as:

$$\|\frac{f}{z}b_{\overline{de}}(P_o + r_{\overline{od}}(P_o, z)) - \frac{f}{z}b_{\overline{de}}(P_o)\|_2 < \frac{f}{z}\beta\|r_{\overline{od}}(P_o, z)\|_2.$$

We then apply our observation of $r_{\overline{od}}$ as:

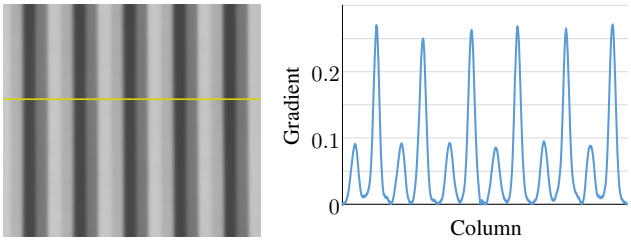
$$\|\frac{f}{z}b_{\overline{de}}(P_o + r_{\overline{od}}(P_o, z)) - \frac{f}{z}b_{\overline{de}}(P_o)\|_2 < \frac{f^2}{z^2}\alpha\beta.$$

Since f, α, and β are known, we can find out the value of z so that the bound becomes less than one: $\frac{f^2}{z^2}\alpha\beta < 1$, s.t. $\forall\ z > 410\,\text{mm}$. This bound enables us to replace the depth-dependent term in Equation (1) with the depth-invariant term as follows:

$$\frac{f}{z}b_{\overline{de}}\left(P_o + r_{\overline{od}}(P_o, z)\right) \approx \frac{f}{z}b_{\overline{de}}\left(P_o\right).$$

3. Calibrating Uneven Double Refraction

We propose a simple yet effective calibration method to obtain the unevenness of the double refraction: the intensity proportion τ between o-ray and e-ray. To this end, we captured stripe patterns printed on a sheet of white paper and computed the gradient magnitudes of the stripe patterns along the disparity direction as shown in Figure 1. The peaks with lower gradient magnitudes correspond to e-ray and higher gradient magnitudes come from o-ray as we set the linear polarizer angle to attenuate e-ray. We compute the intensity proportion τ for these classified peaks (Figure 1). τ is computed only for the pixels near peaks and averaged to a single value. Note that we also calibrate camera parameters and birefringent properties of calcite, following Zhang [9] and Baek et al. [1]. We choose the mean value of the computed τ values for reconstruction.



(a) Filtered stripes (b) One sample line

Figure 1. Calibration of uneven double refraction. (a) We captured the stripe pattern and (b) identify the gradient peaks of o-ray and e-ray based on intensity differences. (c) We estimate the unevenness τ from the proportion between these peaks.

4. Depth Densification and Applications

For many graphics applications, such as augmented reality and depth image-based rendering, it is desirable to obtain a dense depth map instead of a sparse one. Our joint depth-and-color reconstruction algorithm described in the main paper produces sparse depth estimates over edges only. We therefore used the Fast Bilateral Solver [2] with the parameters σ_{pos}, σ_y and σ_{uv} set to 30, 10 and 10 for our image editing applications of the refocusing in the main paper and depth-based segmentation and composition shown in Figure 2. Note that, even though the densification method [2] has been effectively utilized for the image-editing applications, the resulting dense depth maps suffer from lower accuracy on propagated region.

Table 2 shows an example affected by those errors where the performance of 3D object detection degrades when using densified depth maps instead of sparse ones. Developing a robust propagation method remains as a future work.

	Table	Desk	Chair	mAP
Trained with our sparse depth	0.79	0.73	0.67	0.74
Trained with densified depth	0.58	0.33	0.60	0.50

Table 2. 3D object detection AP trained with sparse and dense depth maps and the ground truth dense depth, using the AP metric [5].

Figure 2. Additional editing applications using our estimated color and depth. (a) We perform image segmentation using grab-cut [6] on our depth map to segment and recolor middle owl object. (b) Estimated depth can also be used for depth-aware image composition of the owl.

5. Evaluation

Depth Accuracy with Varying Sparsity As described in the main manuscript, we generated a synthetic dataset of 23 inputs with uneven double refraction. To evaluate our algorithm, we vary the threshold parameters of the validity map to provide different sparsity and accuracy results. Figure 3 shows that accepting larger areas (thus lowering sparsity) leads to a higher average RMSE error while having very high thresholds (high sparsity) can also have a negative im-

pact due to quantization.

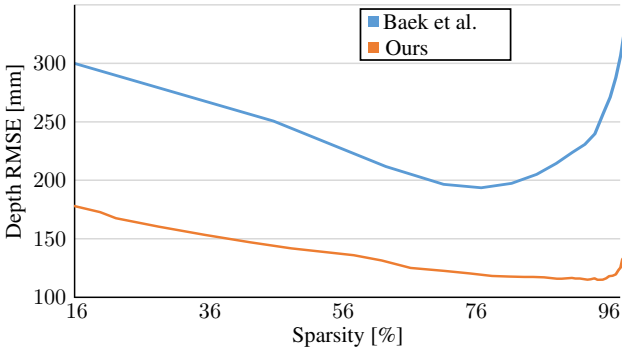


Figure 3. Depth error with varying sparsity on the synthetic dataset. Our method provides significantly lower depth error than Baek et al.'s [1] for all sparsity values.

Chromatic Aberration The chromatic aberration of the calcite crystal has a negligible impact within the range of visible wavelengths. According to the reference [4], the birefringence (the difference of the refractive indices of the e-ray and o-ray) of the medium is *consistent* as 0.17 within the visible wavelengths from 400 nm to 700 nm, .

Impact of Noise Robustness against image noise is important for our system because we rely on uneven double refraction visible around image edges. To evaluate the impact of noise on our method, we used our prototype to capture the panel scene at different noise levels by varying the sensor gain and the shutter time while ensuring the same exposure. The performance of our approach degrades as noise level increases: the depth RMSE rises from 47 mm at 0 db to 148 mm at 24db. However, we found that the impact of noise can be mitigated by applying a simple denoising technique, such as Gaussian filtering: the RMSE falls to 55 mm. Figure 4 presents reconstructed images and depth maps of the panel scene with high noise and results after Gaussian filtering.

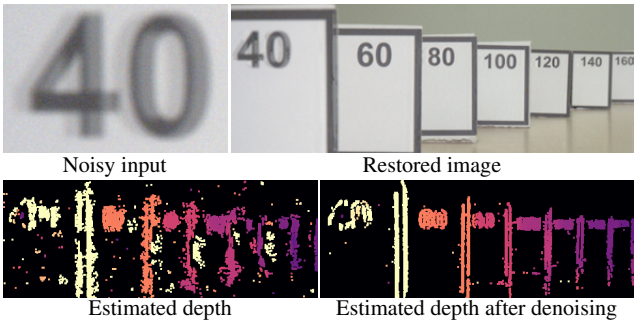


Figure 4. Qualitative evaluation of the impact of noise to evaluate the robustness of our algorithm. We used the gain parameter of 24db. The noise impact can be mitigated by applying a simple Gaussian filtering.

6. Comparison

Color Restoration Methods Our method jointly estimates depth and color efficiently taking 16 ms which is even faster than previous color-only reconstruction methods requiring additional computation in order to estimate depth. In addition to being the fastest restoration method, Figure 5 and Table 3 show that our method achieves state of the art restoration results in terms of accuracy.

	Shih et al.	Wiener	Lucy	Yano et al.	Ours
Color PSNR (dB)	34.20	34.81	33.43	**34.82**	**34.82**

Table 3. Quantitative comparison between our restoration and other methods for our panel scene.

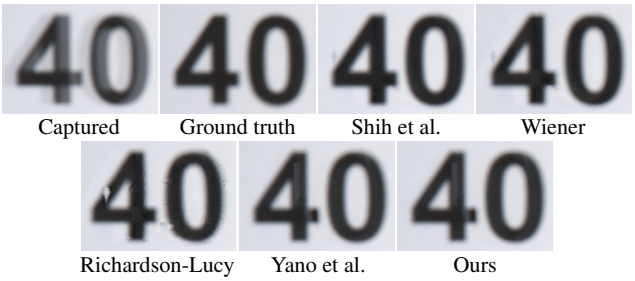


Figure 5. Qualitative comparison between our restoration and other methods for our panel scene.

Learned Depth Estimation Figure 6 shows results with our panel scene using a learned depth prediction method [8]. We used our ground truth restoration shown in Figure 5 as inference input. As the statistical characteristics of the captured scene and the camera parameters are very different from those of the training set, the learning algorithm provides poor results compared to our method, which estimates depth accurately.

Depth from dual pixel Dual pixel phone cameras have been successfully employed to obtain close-range depth [7, 3]. However, their disparity is bounded by the defocus blur, thus not exceeding a few pixels. Figure 7 shows that the disparity that would be obtained through a dual pixel sensor becomes saturated much more quickly than the double refraction disparity does when the depth value increases. We assumed the same setup parameters as ours and DP data downsampled $2\times$ horizontally and $4\times$ vertically as Wadhwa et al. [7] used. We also based our computations on their model:

$$r \approx \alpha L f \left(\frac{1}{g} - \frac{1}{z} \right), \tag{2}$$

Where r is the disparity, α is a linear coefficient, L is the aperture and $g = 800$ mm is the focal plane distance.

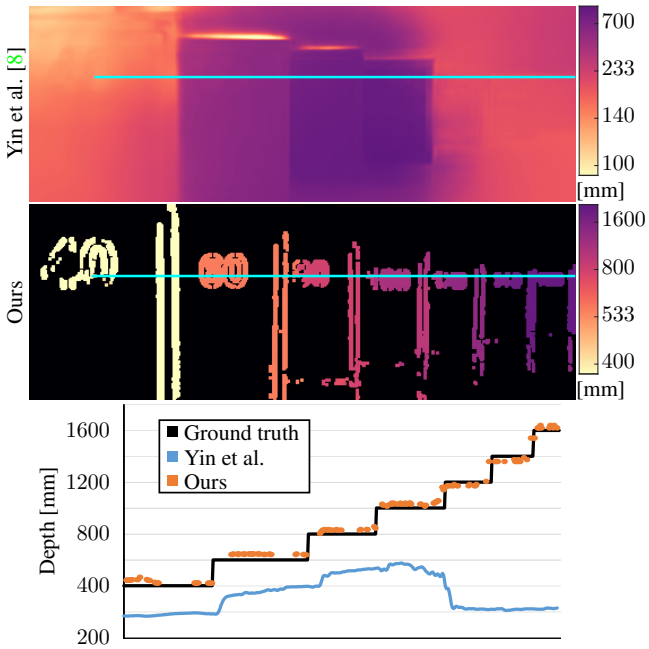


Figure 6. Comparison with depth prediction [8]. The graph shows the depth values along the light blue lines. This method provide a poor depth estimation for this scene. In particular, although the relative depth is well estimated for the first three panels, the absolute values are significantly off.

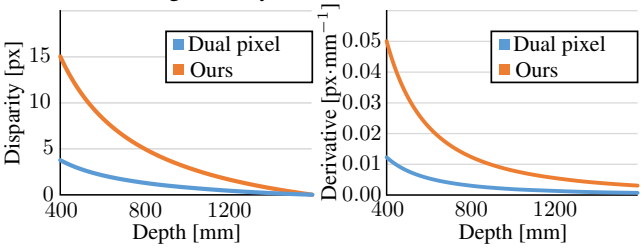


Figure 7. Disparity and absolute disparity's derivative w.r.t the depth against the depth for our setup and for a dual pixel sensor. Both disparity curves are shifted to meet at 0 px at 1600 mm. The disparity's derivative shows the depth discriminative capability, it is much higher for double refraction.

Depth from light-field We compare our method with a light-field method using a plenoptic camera, Lytro Illum as shown in Figure 8. The light-field camera sacrifices spatial resolution for angular resolution to estimate depth information; depth results are competitive with ours. Therefore, the restored image suffers from low spatial resolution when obtaining data on light field (compare the restored results in the middle column).

Depth from Double Refraction As described in the main paper, our method outperforms Baek et al. [1]. Figure 10 and Table 4 show the additional comparison on synthetic scenes.

The dense depth map was produced by the Lytro software.

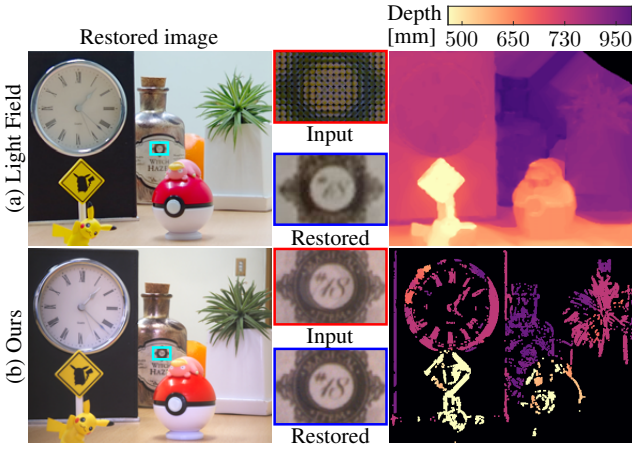


Figure 8. The light-field camera (a) sacrifices spatial resolution for angular resolution to estimate depth information.

	color PSNR (dB)		Depth RMSE (mm)	
Index	Baek et al.	Ours	Baek et al.	Ours
1	24.70	36.10	207	108
2	23.91	37.32	140	101
3	24.77	37.13	259	83
4	25.92	36.90	145	82
5	28.82	36.63	329	146
6	27.77	37.37	142	78
7	29.72	37.31	293	158
8	27.52	36.06	224	63
9	29.35	37.17	327	231
10	27.25	36.28	223	131
11	26.93	36.42	326	228
12	23.51	34.34	148	70
13	26.91	36.44	191	96
14	26.38	35.50	201	50
15	27.30	37.10	179	137
16	28.21	37.46	344	96
17	27.25	37.63	173	139
18	25.99	35.52	311	172
19	24.09	35.88	168	106
20	23.47	35.74	168	95
21	29.12	37.33	136	71
22	28.60	37.55	125	98
23	26.74	37.23	153	124
Mean	26.71	**36.63**	212	**116**

Table 4. Quantitative comparison between Baek et al. and our method on all frames of the synthetic dataset.

7. Additional Results

Figures 11, 12 and 9 show additional results of our system on real and synthetic scenes.

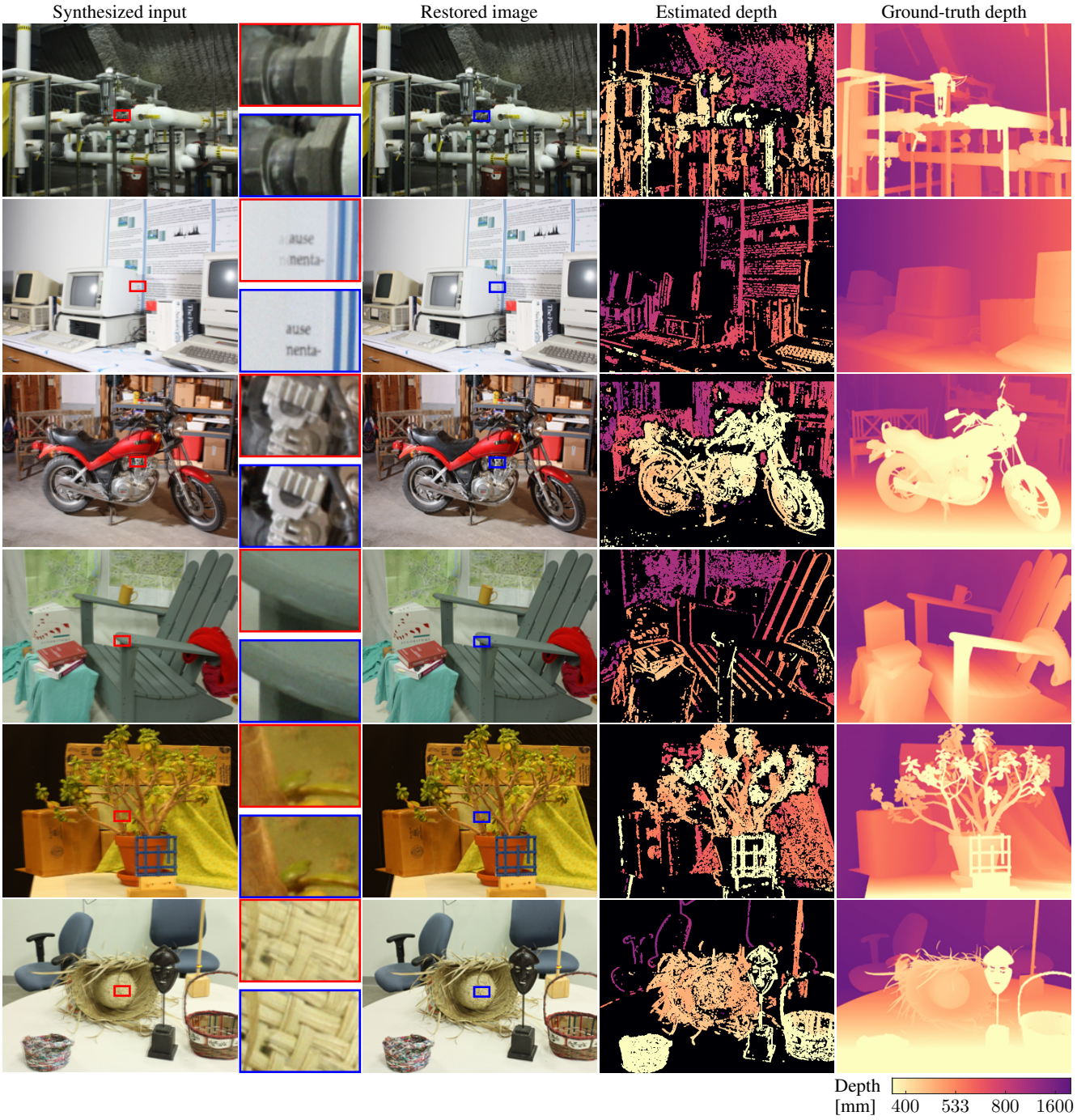


Figure 9. Synthetic results of our algorithm with the ground truth.

References

[1] Seung-Hwan Baek, Diego Gutierrez, and Min H Kim. Birefractive stereo imaging for single-shot depth acquisition. *ACM Trans. Graphics (Proc. SIGGRAPH Asia 2016)*, 35(6):194, 2016. 2, 3, 4, 6

[2] Jonathan T Barron and Ben Poole. The fast bilateral solver. *Proc. European Conference on Computer Vision (ECCV) 2016*, 2016. 2

[3] Rahul Garg, Neal Wadhwa, Sameer Ansari, and Jonathan T. Barron. Learning single camera depth estimation using dual-pixels, 2019. 3

[4] Gorachand Ghosh. Dispersion-equation coefficients for the refractive index and birefringence of calcite and quartz crystals. *Optics communications*, 163(1-3):95–102, 1999. 3

[5] Charles R Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J

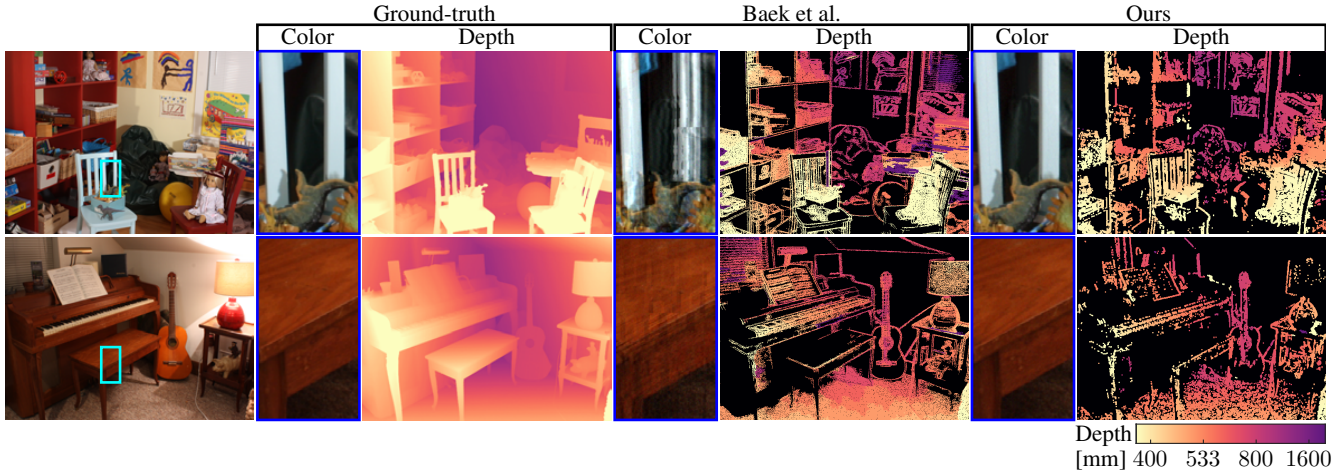


Figure 10. Synthetic results of our algorithm with the ground truth and comparison with [1].

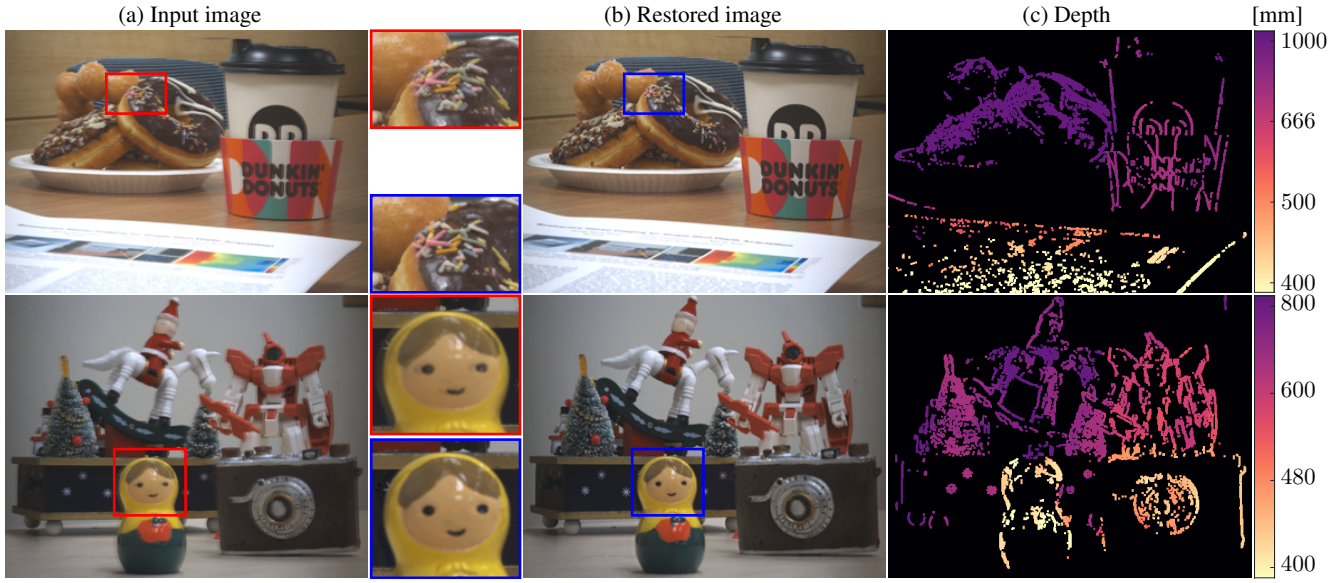


Figure 11. Results with our prototype. (a) Input image shows uneven double refraction. (b) Restoration removes double refraction from input. (c) Sparse depth map.

Figure 12. Limitations of our algorithm. Our method presents suboptimal performance to estimate depth for surfaces with specular highlights or blurred by defocus.

Guibas. Frustum pointnets for 3d object detection from rgb-d data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 918–927, 2018. 2

[6] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph. (TOG)*, 23:309–314, July 2004. 2

[7] Neal Wadhwa, Marc Levoy, Rahul Garg, David E. Jacobs,

Bryan E. Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T. Barron, and Yael Pritch. Synthetic depth-of-field with a single-camera mobile phone. *ACM Transactions on Graphics*, 37(4):113, Jul 2018. 3

[8] Wei Yin, Yifan Liu, Chunhua Shen, and Youliang Yan. Enforcing geometric constraints of virtual normal for depth prediction. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 3, 4

[9] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22, 2000. 2