

Optics Letters

Event-based dual photography for transparent scene reconstruction

XIAOMENG LIU,^{1,*} JOSHUA D. REGO,² SUREN JAYASURIYA,^{2,3} AND SANJEEV J. KOPPAL¹

¹Department of Electrical and Computer Engineering, University of Florida, Florida, USA

²School of Electrical, Computer and Energy Engineering, Arizona State University, Arizona, USA

³School of Arts, Media and Engineering, Arizona State University, Arizona, USA

*Corresponding author: liuxiaomeng51@ufl.edu

Received 15 December 2022; revised 19 January 2023; accepted 19 January 2023; posted 20 January 2023; published 28 February 2023

Light transport contains all light information between a light source and an image sensor. As an important application of light transport, dual photography has been a popular research topic, but it is challenged by long acquisition time, low signal-to-noise ratio, and the storage or processing of a large number of measurements. In this Letter, we propose a novel hardware setup that combines a flying-spot micro-electro mechanical system (MEMS) modulated projector with an event camera to implement dual photography for 3D scanning in both line-of-sight (LoS) and non-line-of-sight (NLoS) scenes with a transparent object. In particular, we achieved depth extraction from the LoS scenes and 3D reconstruction of the object in a NLoS scene using event light transport. © 2023 Optica Publishing Group

<https://doi.org/10.1364/OL.483047>

Introduction. Light transport represents the complex interactions of light in a scene and its capture enables numerous applications in computer vision and graphics [1,2]. The light transport matrix $\mathbf{T}$ describing the linear relation between a light source and a sensor [1,3] maps the programmable illumination $\mathbf{p}$ as a vector to the camera image $\mathbf{c}$ (also vectorized) by the equation $\mathbf{c} = \mathbf{T}\mathbf{p}$. A dual image can be obtained by $\mathbf{T}^T\mathbf{c}$, as viewing from the projector view. However, dual photography has been limited by such challenges as data size and computation.

Our idea is to solve some of these challenges using event cameras. Event cameras have high dynamic range (HDR) and high temporal resolution, and enable a large amount of applications across many fields [4]. Instead of capturing the whole resolution image, as a traditional camera does, the event camera saves the light intensity changes, which helps reduce data capture. A projector–event camera vision system is suggested, to quickly capture light transport and render dual images.

In this Letter, we present a novel hardware setup combining a micro-electro mechanical system (MEMS) flying-spot projector with an event camera. Direct-global event separation using epipolar masks [5] has been applied to the captured light transport for dual applications. In particular, we showcase the HDR advantage of event cameras with respect to a class of traditionally challenging scenes, transparent objects. Our contributions can be summarized as follows:

- A novel projector-camera system to capture light transport using a flying-spot projector and an event camera.
- Imaging and depth estimation of event light transport of a line-of-sight (LoS) scene having a transparent object.
- Silhouette detection and 3D shape estimation of transparent objects behind a barrier, using event light transport of a rotating non-line-of-sight (NLoS) scene.

Related work: efficient capture of light transport. Capturing the columns of the light transport matrix is done by illuminating a single region with a focused light source (e.g. flying-spot, single projector pixel). However, commercial projectors produce low signal-to-noise ratio (SNR) with single pixel illumination; thus, a number of coded illuminations with demultiplexing have been proposed to increase the SNR [6]. For fast capture of light transport, Henderson *et al.* proposed a setup [7] to improve the “flying-spot scanner” by using a beam of light reflected on a programmable MEMS mirror. Liu *et al.* [8] recently further improved the light efficiency of this setup by replacing the LED light source with a powerful RGB modulated white laser. Similarly, we leverage a flying-spot projector with a laser source for our hardware setup.

Data storage efficiency. Light transport capture results in a number of measurements, requiring large amounts of storage and a long processing time. To speed up the acquisition and reduce measurements, Sen *et al.* [2] used adaptive illumination to capture the light transport and generate dual photography, describing a full reflectance field (~66 TB) with only 144 HDR images. Peers *et al.* [9] utilized compressive sensing via non-adaptive illumination patterns, which reduced the measurements by two orders of magnitude compared with brute-force acquisition. Sen and Darabi [10] extended compressive sensing to dual photography. Although this method reduced the number of measurements, it still required 600 full camera resolution images (648×488) to be stored. In contrast, our method only requires the event camera raw file to be stored to reconstruct the light transport, improving the storage efficiency, and reducing the capture and preprocessing time.

Transparent object reconstruction. Transparent object depth recovery using RGB-D cameras includes deep-learning with transparent object segmentation [11] and local implicit neural representation on ray–voxel pairs [11], as well as depth

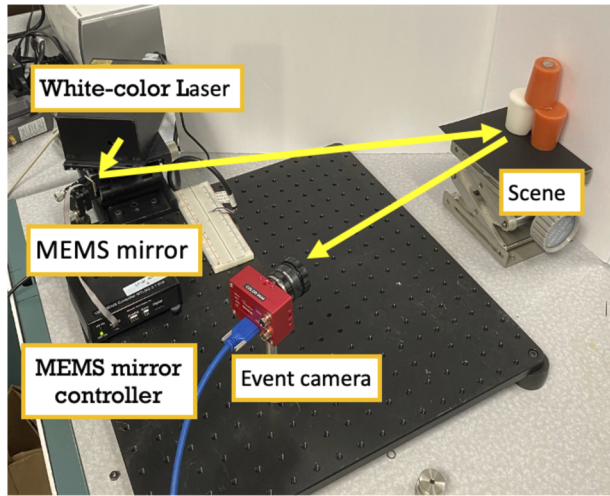


Fig. 1. Hardware setup.

with silhouette information [12]. Wu *et al.* [13] used multi-view RGB capture to optimize a model with constraints on light refraction paths and silhouettes. To our knowledge, no one has shown transparent object reconstruction for dual photography or occluded scenes.

Event-based dual photography: hardware setup. The acquisition speed of light transport is influenced by sensor capture speed, data storage, and processing. Our choice of MEMS flying-spot projector and event camera achieves a satisfactory trade-off in this area.

To capture the light transport for dynamic scenes, flying-spot scanners utilize MEMS mirror-based projectors coupled with high-speed cameras [7,8]. The light efficiency of a flying-spot projector is at least a factor of m larger than a conventional projector, where m denotes the speed advantage of a conventional projector over flying-spot projectors [8]. For these reasons, we adopted a flying-spot scanner design for our illumination.

Figure 1 shows the hardware setup of our projector–event camera system. The projector in our system is similar to the flying-dot projector of recent work [8]. A programmable MEMS modulated mirror moves a beam of white light across the scene. The light is emitted from a laser whose dot size is 1 mm in diameter. The MEMS mirror coupled with it is 5 mm in diameter. The moving path of the light is a 2D Lissajous pattern with 50 horizontal lines.

We used a DAVIS346 color event camera with a resolution of 346×260 , combined with an Edmund 8 mm MegaPixel fixed focal length lens to capture the light changes as the light spot moves over the scene. The raw data generated from the event camera were the events, including timestamps, x and y coordinates, and polarities of light changes.

A single dot image refers to an image captured when a bright dot illuminates the scene. The event data stream is separated by the timestamp into 29- μ s time windows and then accumulated to form single dot images. The single dot images for one complete raster scan are vectored to the columns of the light transport matrix. One scan of a scene takes approximately 3 min to complete. Dual imaging for both LoS and NLoS scenes is based on these images.

Storage. The required storage for a light transport is governed by $S = R \times N_T$, where S is the data size, R is the sensor resolution, and N_T is the number of light transport columns. For an

event camera: $S_{\text{evt}} = \sum_{k=1}^{N_T} R_k$, where R_k refers to active pixels for the k th column of light transport. The parameter R_k is scene-dependent and is defined as $R_k = \alpha_k \times R_{\text{evt}}$, where α_k is the scene coefficient and R_{evt} is the full event camera resolution. Assuming, on average, that β of the pixels have changed intensity, the data storage can be $S_{\text{evt}} = \beta \times R_{\text{evt}} \times N_T$.

We compared an event camera with a RGB camera, $S_{\text{evt}} : S_{\text{RGB}} = \beta \times R_{\text{evt}} \times N_T : R_{\text{RGB}} \times N_T$, and concluded that event cameras required only β of the storage of RGB cameras.

We compare our method with prior work [8], which used a RGB high-speed camera. Our event camera has a resolution of 346×260 . The event light transport has 7120 columns, requiring 0.5 GB storage space. The high-speed camera had a resolution of 328×768 . It captured a light transport having 9980 columns, resulting in 15 GB data. So we have an efficiency factor of $\beta = 0.013$ for the experiments in this Letter.

Dual imaging of transparent objects. In this section, we present experiments using the novel hardware setup in Fig. 1. This was, to our knowledge, the first time that an event camera was involved in capturing light transport. We calibrated the camera–projector system with a pre-captured Lambertian plane, a similar method to Liu *et al.* [8]. Homography information was extracted and the epipolar lines on the camera image plane were computed. Epipolar and non-epipolar separation were applied to the LoS light transport [14]. Dual imaging for both LoS and NLoS scenes utilized direct-global event separation with epipolar masks [5].

Dual depth validation. Figure 2 presents the validation of our dual depth imaging method. Two objects were placed in the scene, at different distances from the projector. We accumulated the events within a certain time window to form the event dot image, defined as I . We used P and S to represent the dot images of the Lambertian plane and the scene, respectively. Algorithm 1 shows the depth computation process. A direct-global separation method similar to that of O’Toole *et al.* [5] was applied to event dot images, to obtain the direct and global components. We

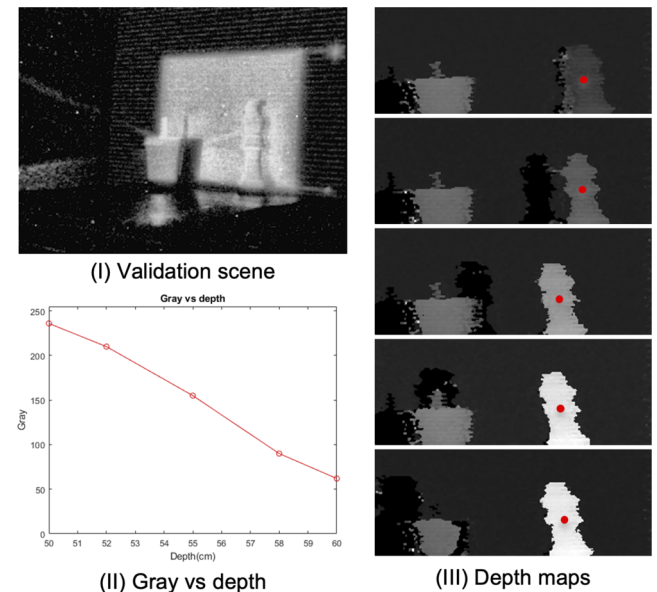


Fig. 2. Depth validation for opaque scene: (I) example of validation scene; (II) gray as a function of depth for the validation; (III) depth maps of two opaque objects placed at different distances. The red points were used to generate the graph in (II).

Algorithm 1. Generating LoS dual depth

- 1: **Input:** Light transport from single scan for both Lambertian plane and LoS scene (i.e., accumulated event dot images from event camera)
- 2: **Output:** Dual depth map of LoS scene
- 3: **for** each dot image for plane P_i and scene S_i **do**
- 4: $P_{\text{direct}}^i, P_{\text{global}}^i = \text{separate}(P_i)$ ▷ Epipolar separation
- 5: $S_{\text{direct}}^i, S_{\text{global}}^i = \text{separate}(S_i)$
- 6: $c_{\text{plane}}^i = \text{findCenter}(P_{\text{direct}}^i)$
- 7: $c_{\text{scene}}^i = \text{findCenter}(S_{\text{direct}}^i)$
- 8: $d_i = \frac{\alpha}{\|c_{\text{plane}}^i - c_{\text{scene}}^i\|}$ ▷ This is depth
- 9: **for** each pixel (i, j) on dual map **do**
- 10: Index = homography(i, j)
- 11: map(i, j) = d_{Index}

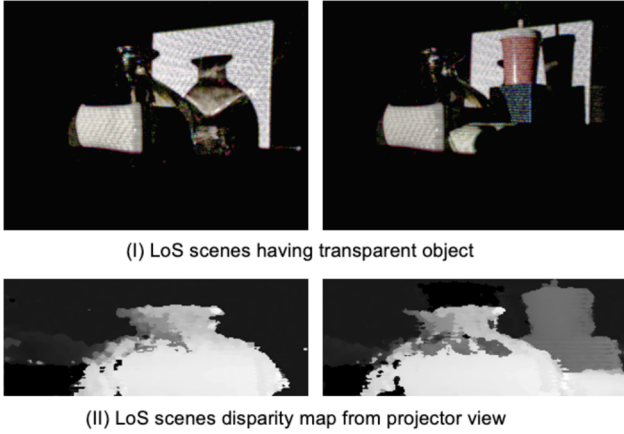


Fig. 3. LoS dual results: (I) LoS scenes of a glass bottle with milk; (II) dual depth maps from projector view of the scenes in (I). The transparent glass has been detected.

assumed that the dot had the highest brightness, so the centroid of the biggest region of an event dot image was considered as the center of dot C . The homography information mentioned in Algorithm 2 was obtained in a calibration process, similar to that of Henderson *et al.* [7]. The validation results are shown in Fig. 2(III).

Dual depth for LoS scenes using direct event data. The dual algorithm to extract depth information is shown in Algorithm 1. For a LoS scene having a transparent object, we generated the dual depth image in Fig. 3. Capturing transparent objects has been difficult because the majority of the light penetrates the object. Our method works for two reasons. First, the weak reflection on the object surface is captured as events in the event camera because of its HDR. Second, this weak reflection is separated cleanly, since we use the direct component of the light transport.

Separation for transparent objects. The epipolar separation method [5] has limitations when splitting the light located on the epipolar plane. This method considered all light in the epipolar plane as the direct component. However, strong global components including reflections can become located on the epipolar plane, and can be considered direct components by mistake. To correct this, we generated the floodlit results for both the scene and the Lambertian plane. Then we subtracted the Lambertian plane from the scene and conducted image segmentation. We

created a binary mask, based on the segmentation result, which only selected the object region. According to the homography information, which mapped the dual pixel values to single dot images from the camera view, we applied the object mask to the epipolar separation results, the direct component. Through this additional masking process, we can obtain the fixed direct component and rough estimated dual depth image shown in Fig. 3. *Note that the glass is invisible in the floodlit results but reconstructed in our 3D estimate.*

Dual NLoS 3D reconstruction using global event data. To reconstruct an occluded transparent object, we first create a silhouette edge, which we call a binary mask. Event cameras have a HDR that helps capture this silhouette edge in a way that traditional RGB cameras cannot match, the full range of the signal that traditional RGB cameras cannot capture. Figures 4(III) and 4(IV) show the dual binary masks (silhouette edges) generated from RGB and event sensor data using a method we describe next. For a NLoS scene, the event sensor captured more signal than the RGB sensor, and thus formed a dual binary mask with more detail. We now explain how to use several of these masks for a rotating scene to reconstruct a transparent object behind a barrier.

Our algorithm is shown in Algorithm 2. The key idea is to detect when the laser dot hits the transparent object (behind the barrier) by measuring a significant increase in global illumination (in front of the barrier). For transparent objects, this global illumination is a refractive caustic. We threshold the events to detect an increase in global illumination. We studied the standard deviations of the dot images and chose 0.02, which was the upper boundary of the low-signal images.

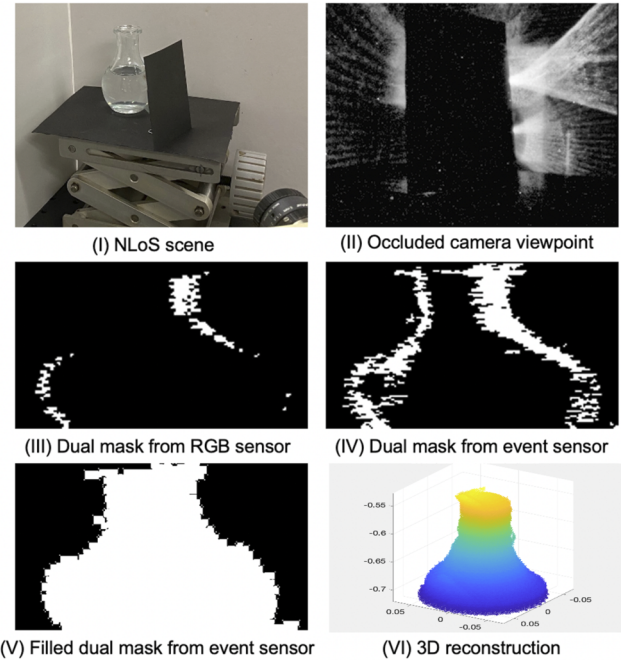


Fig. 4. 3D reconstruction of NLoS scene: (I) scene image.; (II) scene floodlit image from event camera view; (III) dual mask using RGB sensor data; (IV) dual mask using event sensor data; (V) filled binary mask from (IV); (VI) 3D reconstruction result. The result was obtained by feeding 20 duplications of (V) to the SpaceCarving algorithm [17]. Both the height/width and height/depth ratios of the rotationally symmetric object were 1.30; our reconstruction showed ratios of 1.35.

Algorithm 2. Generating NLoS dual mask

```

1: Input: Light transport from single scan (i.e., accumulated event
   dot images from event camera)
2: Output: Dual binary mask of NLoS scene
3: for each dot image  $I_i$  do
4:    $I_{\text{direct}}^i, I_{\text{global}}^i = \text{separate}(I_i)$ 
5:    $\sigma(i) = \text{std}(I_{\text{global}}^i)$ 
6: ObjectDotImages = find( $\sigma_i > \text{cutoff}$ )
7: for each pixel  $(i, j)$  on dual mask do
8:   Index = homography( $i, j$ )
9:   if Index  $\in$  ObjectDotImages then
10:    mask( $i, j$ ) = 1
11:   else mask( $i, j$ ) = 0

```

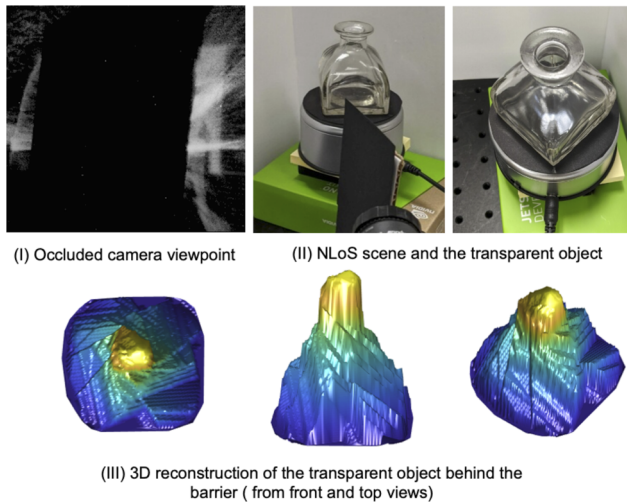


Fig. 5. NLoS dual results. (I) NLoS scene floodlit from event data. (II) NLoS scene setup with black barrier standing between camera and object. The transparent object has a square base. (III) Top 3D reconstruction of object from 18 frames of NLoS scene when object rotates.

The SpaceCarving algorithm [15–18] can create a 3D reconstruction from images captured from different viewing angles. We fill silhouette edges with a simple version of the watershed algorithm, as from Fig. 4(IV) to Fig. 4(V). For our first result in Fig. 4(VI), we duplicated the filled binary mask [Fig. 4(V)], simulating a rotation of a rotationally symmetric object. We also captured 18 frames of a rotationally asymmetric NLoS scenes in Fig. 5. Sending these dual binary masks to the SpaceCarving algorithm, we 3D reconstructed both transparent objects from the projector view, as shown in Figs. 4 and 5.

Reconstruction accuracy. To the best of our knowledge, ours is the only method that can reconstruct transparent objects behind a barrier. Figures 4(II) and 5(I) show the occluded scene from the point-of-view of the camera, and the object cannot be seen directly at all. While there are errors in the reconstruction (e.g., the neck of the glass), the overall accuracy is good, and the physical height-to-width and height-to-depth ratios of the object in Fig. 4 were 1.30 and 1.30, respectively, compared with our reconstructions of 1.35 and 1.35. Qualitatively, Fig. 5 also

shows the square base of the glass.

Please see accompanying [Visualization 1](#).

Conclusion. To achieve lightweight light transport capture, we combined a flying-spot projector and an event camera that improves data efficiency and low SNR issues. We also captured LoS and NLoS light transport slices (namely, direct and global separation) of a transparent object. We present two applications in dual imaging: depth mapping and 3D reconstruction for transparent objects. Further, we share our event data and codes from this hardware setup to the public for further research.

Funding. National Science Foundation (1909192, 1909729, 1942444).

Acknowledgments. The authors thank the anonymous reviewers and the editors for their helpful feedback.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in [19].

REFERENCES

1. R. Ng, R. Ramamoorthi, and P. Hanrahan, *ACM Trans. Graph.* **22**, 376 (2003).
2. P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. Lensch, in *ACM SIGGRAPH 2005* (ACM, 2005), pp. 745–755.
3. J. T. Kajiya, in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques* (ACM, 1986) pp. 143–150.
4. G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, and K. Daniilidis, *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 154 (2022).
5. M. O'Toole, J. Mather, and K. N. Kutulakos, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2014) pp. 3246–3253.
6. Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 1339 (2007).
7. K. Henderson, X. Liu, J. Folden, B. Tilmon, S. Jayasuriya, and S. Koppal, *IEEE Trans. Comput. Imaging* **6**, 529 (2020).
8. X. Liu, K. Henderson, J. Rego, S. Jayasuriya, and S. Koppal, *Opt. Express* **29**, 18362 (2021).
9. P. Peers, D. K. Mahajan, B. Lamond, A. Ghosh, W. Matusik, R. Ramamoorthi, and P. Debevec, *ACM Trans. Graph.* **28**, 3 (2009).
10. P. Sen and S. Darabi, *Computer Graphics Forum* **28**, 609 (2009).
11. S. Sajjan, M. Moore, M. Pan, G. Nagaraja, J. Lee, A. Zeng, and S. Song, 2020 IEEE International Conference on Robotics and Automation (ICRA) (IEEE, 2020) pp. 3634–3642.
12. Y. Ji, Q. Xia, and Z. Zhang, *Int. J. Opt.* **2017**, 9796127 (2017).
13. B. Wu, Y. Zhou, Y. Qian, M. Gong, and H. Huang, "Full 3D reconstruction of transparent objects," *arXiv arXiv:1805.03482* (2018).
14. T. Oo, H. Kawasaki, Y. Ohsawa, and K. Ikeuchi, in *Asian Conference on Computer Vision* (Springer, 2006) pp. 908–917.
15. A. W. Fitzgibbon, G. Cross, and A. Zisserman, *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments* (Springer, 1998), pp. 155–170.
16. M. Zins, "Spacecarving," GitHub (2019) [accessed 24 February 2023] <https://github.com/zinsmatt/SpaceCarving>.
17. K. N. Kutulakos and S. M. Seitz, *International journal of computer vision* **38**, 199 (2000).
18. L. Davis, ed., *Foundations of Image Understanding*, (University of Maryland, 2001).
19. X. Liu, J. D. Rego, S. Jayasuriya, and S. J. Koppal, "Event-based dual photography for transparent scene reconstruction," *Focus Lab* (2023), <https://focus.ece.ufl.edu/event-based-dual-photography-for-transparent-scene-reconstruction/>.