

Imaging Behind Occluders Using Two-Bounce Light

Connor Henley, Tomohiro Maeda, Tristan Swedish, and Ramesh Raskar

Massachusetts Institute of Technology, Cambridge MA 02139, USA
`{co24401, tomotomo, tswedish, raskar}@mit.edu`

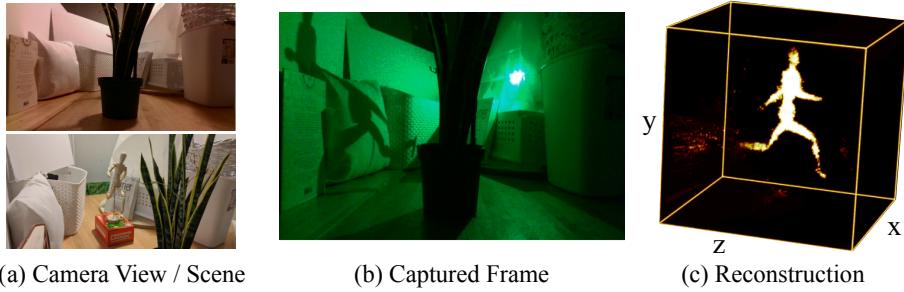


Fig. 1. Imaging behind occluders using visible surfaces on opposing sides of a hidden space. (a) Top: A mannequin is hidden from the camera. Bottom: A third person view of the occluding plant. (b) We illuminate points lying to one side of the hidden space, and observe distorted shadows that hidden objects carve out of the two-bounce light signal, which scatters off of the opposing visible surface towards our camera. (c) Euclidean 3D reconstruction of the mannequin.

Abstract. We introduce the new non-line-of-sight imaging problem of *imaging behind an occluder*. The behind-an-occluder problem can be solved if the hidden space is flanked by opposing visible surfaces. We illuminate one surface and observe light that scatters off of the opposing surface after traveling through the hidden space. Hidden objects attenuate light that passes through the hidden space, leaving an observable signature that can be used to reconstruct their shape. Our method uses a simple capture setup—we use an eye-safe laser pointer as a light source and off-the-shelf RGB or RGB-D cameras to estimate the geometry of relay surfaces and observe two-bounce light. We analyze the photometric and geometric challenges of this new imaging problem, and develop a robust method that produces high-quality 3D reconstructions in uncontrolled settings where relay surfaces may be non-planar.

Keywords: Non-line-of-sight imaging, Computational Photography

1 Introduction

Traditional optical imaging techniques produce images using measurements of light that has propagated directly, along a straight and unoccluded line of sight, from an object or scene of interest to one's imaging sensor. Non-line-of-sight (NLOS) imaging techniques, by contrast, generate images from measurements of light that has traveled from the object or scene of interest via indirect paths that typically include reflections off of intermediate surfaces. NLOS imaging techniques are particularly useful for looking behind things—walls, buildings, vehicles—any opaque surface that blocks one's line of sight to objects at greater depths.

In this paper, we introduce a new NLOS imaging problem, which we will refer to as the problem of *imaging behind an occluder*. The problem can be described as follows: any opaque occluder will block an observer's line of sight to all surfaces that lie behind it, creating a hidden volume that extends to infinite depth behind the occluder. In some scenarios, an observer will have an unobstructed view of surfaces that lie on either side of this hidden volume. This scenario is depicted in Figure 3. Such a geometry immediately suggests an elementary measurement: the observer can illuminate a point lying to one side of the hidden volume, and observe a point lying on the other side. The line segment that connects these two points passes through the hidden space. If no occluding surfaces lie on this segment, then light reflected off of the illuminated point will propagate to the observed point, which will then also appear illuminated. On the other hand, if an occluding surface does lie upon this line segment, then the observed point will lie in that surface's shadow, and appear dim. The problem of imaging behind an occluder is the problem of interpreting these shadows to produce a 3D reconstruction of the hidden scene.

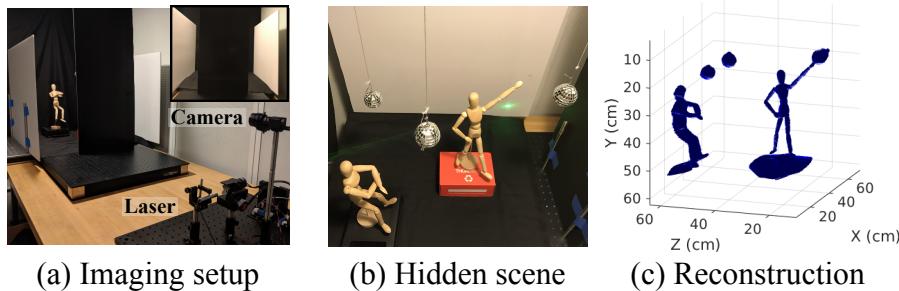


Fig. 2. Our method produces accurate reconstruction of hidden scenes containing multiple objects using a laser pointer and off-the-shelf RGB or RGB-D cameras.

1.1 Contributions

We summarize the contributions of our paper as follows:

1. Formulation of a new non-line-of-sight imaging problem of *imaging behind an occluder*. We discuss challenges and sources of error in *imaging behind an occluder*, particularly in natural settings in which extraction of shadows and 3D estimation of visible surfaces are prone to error
2. Method that solves this problem with a space carving approach that exploits two-bounce light measurements. Ours is the first NLOS method to recover detailed 3D shapes of hidden objects without specialized equipment. We use a laser pointer and off-the-shelf RGB or RGB-D cameras.
3. Reconstruction results that highlight the capability of our method to recover fine structure of hidden objects, capture video reconstructions of moving scenes, and handle non-planar and non-continuous visible surfaces

Behind-an-occluder methods can be applied to a variety of useful tasks, including vision in cluttered environments such as forests, maintaining spatial awareness of the space behind neighboring vehicles on the road, seeing inside buildings during search-and-rescue operations, or determining the 3D shape of the back-facing side of an object. We present additional results that highlight potential applications of our method to autonomous driving and search-and-rescue operations.

2 Related Work

2.1 Non-Line-of-Sight Imaging

Most NLOS imaging methods published to date have attempted to solve the challenging problem of *seeing around the corner*. In the around-the-corner problem, an observer can only learn about the hidden scene from light that is reflected off of intermediate surfaces *after* it has reflected off of or been emitted by the hidden scene itself.

Velten et al. [20, 27] demonstrated the first 3D reconstruction of an object hidden around the corner by using time of flight (ToF) measurements to provide constraints on hidden object locations. Since then, many robust and efficient algorithms to image around corners have been proposed [1, 9, 15, 28]. Others exploit speckle patterns [10], spatial coherence [3], and radiometric (intensity) measurements [6, 11, 24] for NLOS imaging. Around-the-corner methods typically require measurement of three-bounce photons. Our method captures two-bounce photons and does not require ToF measurements to recover the 3D geometry of hidden objects.

Occlusions constrain the set of paths that light can take as it propagates through the hidden scene and back to the observer. *Occlusion-assisted* imaging techniques exploit knowledge of these constraints to produce a forward light transport model that can be inverted to estimate hidden scene properties.

Bouman et al. [5] produce 1D images of hidden scenes by exploiting the occlusions that occur as light propagates past a corner or an edge. Others exploit occlusions that occur within the hidden scene to recover surface albedos [21, 26], and 4D light fields [2] around corners. In these techniques, the occluders and hidden scenes are considered separate, and the former is exploited to learn about the latter. In our method, we directly measure the shadows cast by hidden objects—in effect, the hidden object *is* the occluder.

We refer readers to [16] for a comprehensive overview of NLOS imaging.

2.2 Shape from Shadows

In *shadowgram imaging* methods, the shape of an object is estimated from a series of shadows cast onto a planar screen for point sources placed at various locations [22, 29]. Shadowgram methods are typically applied in controlled, object scanning setups and require the placement of point sources with a direct line of sight to the scene being imaged. Our method is designed to reconstruct scenes that might be entirely hidden from view, and to utilize shadows projected onto visible surfaces that may be non-planar, have varied surface albedo, and have to be estimated by the observer.

Shadowgram methods exploit the fact that a sharp shadow cast onto a plane by an object that is illuminated by a point source can be interpreted as a 2D *silhouette image* of that object, taken from the perspective of the point source. The problem of determining the three-dimensional shape of an object from a series of two-dimensional silhouettes has been studied extensively by the computer vision community [14, 18, 25]. With multiple silhouettes, taken from multiple camera points, an object’s shape can be confined to an intersection of affine cones, referred to as the *visual hull* [13]. The first shape-from-silhouettes (SfS) algorithm is often credited to Baumgart [4]. Later, Martin and Aggarwal [17] introduced a volumetric space carving approach to SfS in which voxels lying outside the visual cone are “carved” away, forming a visual hull.

We modify our space carving approach so that it is robust to the errors in visible surface geometry and shadow classification. Some researchers have explored robust *probabilistic carving* methods. These methods are typically designed for shape-from-silhouettes applications that require a direct view of an imaged scene [8, 12, 23], and cannot be applied directly to our problem of space carving from distorted shadows cast onto arbitrary surfaces by objects that are hidden from view.

3 Imaging Behind Occluders

Consider the flatland scene presented in Figure 3. In this scene, a large fraction of the observer’s field of view is blocked by an occluder. We want to know what is behind that occluder—specifically, we want to find out whether any opaque object is occupying the hidden point \mathbf{x} .

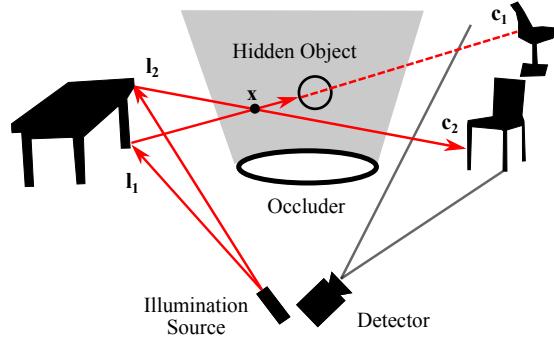


Fig. 3. Visibility Carving Our algorithm reconstructs the shape of hidden objects by determining the set of line segments l_i-c_i that pass through the hidden scene. Here we observe that light can travel from points l_2 to c_2 without occlusion, and thus infer that hidden point x is unoccupied.

3.1 The Elementary Measurement

We illuminate the visible point l_1 , and observe point c_1 , chosen such that the line segment connecting l_1 and c_1 passes through x . In this case, we note that point c_1 lies in shadow. From this measurement, we can infer that a hidden object must lie *somewhere* along the line segment l_1-c_1 , but cannot say for certain that this object is located at point x .

We then take a second measurement, illuminating a different point l_2 and observing point c_2 , again chosen such that the line segment l_2-c_2 passes through x . This time the observed point c_2 appears to be illuminated by light that has been diffusely reflected from the visible surface at l_2 . This measurement is more informative. If light has traveled from l_2 to c_2 without occlusion, then there cannot be an occluding surface at point x or at any point along the segment l_2-c_2 .

This is the elementary measurement that underlies our method for seeing behind occluders. The measurement can be extended to a reconstruction problem by discretizing the hidden space into a grid of points $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ to be probed. Instead of probing the occupancy of each point x_i with a series of $l-c$ pairs, we can efficiently collect this data by using a camera to observe an *area* on the visible surface rather than observing single points. In this case, every measurement probes a large number of hidden points in the scene simultaneously.

3.2 Method

Our method can be decomposed into three basic steps: data capture, pre-processing, and reconstruction.

In the data capture step, we use a galvo-scanned laser to illuminate points on the visible surface. For each laser scan position, we take two photos: one short

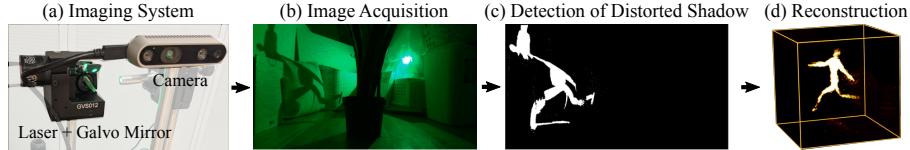


Fig. 4. A pipeline for imaging behind occluders. (a) Our imaging system consists a RGB (or RGB-D) camera and a laser scanned by a two-axis galvo mirror. (b) We use the laser to illuminate points on the visible surfaces and then capture images of the laser spot and shadow. (c) We detect the pixels that belong to the shadowed regions where light is blocked by the hidden object to perform (d) 3D reconstruction.

exposure photo that allows us to pinpoint the location and (if desired) brightness of the laser spot, and a second, longer exposure photo that captures the shadows cast onto the opposing visible surface with high contrast. Pictures of our data capture equipment are shown in Figures 4(a).

In the pre-processing step, we loop through our acquired photo stacks. For each short exposure photo, we pinpoint the image-space position of the laser spot and convert this position to 3D world coordinates. If we are using an RGB camera (as in Section 4) we rely on prior knowledge of visible surface geometry to make this estimate. If geometry information is not available, we directly measure the 3D position using an RGB-D camera. For each longer exposure photo, we select a region of interest that we expect will contain cast shadows that are informative about the hidden scene. We classify each pixel in this region as *shadowed* or *lit* or, in some cases, *unknown*. Our choice of *shadow segmentation* criteria can be made simple (such as a binary threshold) or more complex depending on the complexity of the visible scene. An example of shadow segmentation is shown in Figures 4(b) and (c).

Finally, in the reconstruction step, we discretize the hidden space into a 3D grid of points or voxels. Given an illumination spot \mathbf{l} and its associated shadow image, we project each element of the hidden grid onto the opposing visible surface, using \mathbf{l} as the center of projection. We then determine whether the projection of this element lies *inside* the shadowed region of the visible surface or *outside* of the shadows, in a lit region. It is also possible that the projection does not land on the visible surface at all, or lands on a region of ambiguous classification. If the element is a voxel, it may project to a mixture of shadowed and lit pixels. In this case, we are conservative—classifying the voxel as *outside* only if it projects to all lit pixels, and *inside* only if it projects to all shadowed pixels. We perform this test on each element, for each illumination point and shadow image in our stack. We count how many times an element has been classified as inside and as outside. Once all frames have been processed, we use these counts to determine which elements are occupied and which are empty. In the naive version of our algorithm that is used in Section 4, an element is assessed as empty if it is classified as outside in at least one frame. In Sections

5 and 6 we apply a probabilistic thresholding procedure that is robust to errors in the inside/outside test. An example reconstruction result is shown in Figure 4(d). In this example voxels are colored by probability of occupancy.

4 Reconstructing the Shape of Hidden Objects

In this section, we demonstrate that our method is capable of producing highly detailed reconstructions of stationary hidden scenes, and also demonstrate how our method could be used to capture video reconstructions of hidden objects that are moving. We place hidden objects in a simple testbed consisting of two white relay walls and a black occluding wall. This simple testbed allows us to demonstrate the potential capabilities of our method when errors in pre-processing—that is, visible surface estimation and shadow segmentation errors—are minimal. In later sections, we will address the problem of reconstructing a hidden scene when pre-processing errors are unavoidable.

4.1 Implementation

Our testbed consisted of two white *observation walls*, and a black *occluding wall* placed between the scene-to-be-imaged and all imaging equipment. The observation walls were 61 cm \times 76 cm rectangles oriented parallel to one another and spaced 76 cm apart. A photo of the testbed is shown in Figure 2. We illuminate the two observation walls at a series of points using a green CW laser with a power of \sim 5 mW scanned with a two-axis scanning galvo mirror system, and use a Point Grey Blackfly RGB camera to capture images. To segment shadow images into lit and shadowed regions, we binarize the pixels in a region of interest using either a hand-tuned threshold or a threshold set adaptively using Otsu’s method [19]. We de-noise this binary image using a bilateral filter and then use a second threshold to separate the filtered images into shadowed and illuminated pixels. We found that this second filtering step resulted in more robust pixel classifications.

In this set of collections, we assume that the geometry of the two observation walls is known to the observer. Since the observation walls are planar, this means that we can measure the position of the laser spot in pixel space and then convert that measurement to 3D world coordinates using a homography. Likewise, to test the inside/outside status of a hidden point \mathbf{x} , we find the point at which the ray drawn from \mathbf{l} and through \mathbf{x} intersects the opposing observation wall, and then use a homography to convert this point of intersection to pixel space. We then compare the ray’s point of intersection with the estimated shadow boundaries in pixel space to determine whether it falls inside or outside of the illuminated region. All points that accrue one or more *outside* classifications are carved out of the reconstructed scene. *Inside* classifications are not used.

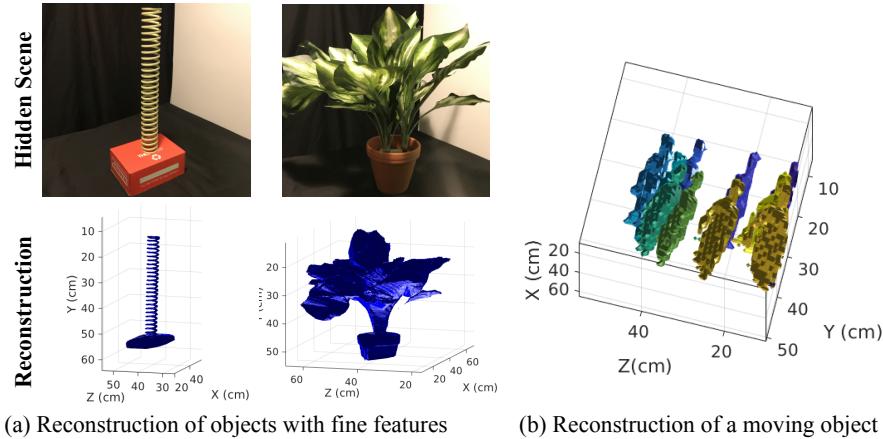


Fig. 5. Reconstruction results. (a) Our method recovers the complex geometry of a spring (left) and a plant (right). An additional reconstruction of a disco dance party can be seen in Figure 2. We use the shadows cast by hidden objects to reconstruct their 3D shape. (b) Snapshots from a 15 FPS video reconstruction of a moving mannequin. Reconstructed frames are colored by time. For clarity, snapshots in this figure are spaced 3 seconds apart. The full video reconstruction can be found in the supplemental material.

4.2 Stationary Hidden Objects

We scanned a variety of stationary objects in this testbed. Photos of three of these objects, along with our estimation of each object’s shape, can be seen in Figures 2 and 5. Our method reproduces the fine structure of reasonably complex objects. On the left side of Figure 5(a), we recover the shape of a spring. Our estimate correctly reconstructs the spring as a singular object, with no breaks along the coil. On the right side of Figure 5(a), we estimate the shape of a plant and are able to resolve individual leaves despite significant self-occlusion. Finally, as shown in Figure 2(a), we recover the shape of a dance party. We believe that the quality of the estimated scene should be more than sufficient for high level tasks such as pose estimation. We also note that the specular surfaces of the scene’s three disco balls had no effect on our ability to recover their shapes.

For each object, we scanned a 9×11 grid of illumination points on both the left and right observation walls. Points on these grids were evenly spaced and spanned almost the entire area of each observation wall. The total acquisition time for each object was 27 seconds. After the acquisition and shadow segmentation, space carving was initiated on a grid of hidden points with a 0.25 cm spacing.

4.3 Moving Hidden Object

By reducing the number of scan points per reconstruction, we can capture videos of dynamic hidden scenes. Although these videos do not capture fine detail, we are able to recover the approximate size, shape, and position of a moving object using only four observed shadows per video frame.

Our reconstructed video can be found in the supplemental material. Snapshots of the result are plotted in Figure 5(b). The hidden object was a mannequin that was slowly moved around the hidden area. We acquire 15 shadows per second, such that the four-point scan pattern can be completed in .27 seconds. Space carving was executed on a 55 cm × 45 cm × 55 cm grid of hidden points, with a grid spacing of 1 cm.

5 Overcoming Geometric and Photometric Challenges of Imaging Behind Occluders

The physical principle that underlies our method of imaging behind occluders using two-bounce light is simple: if we observe that light can travel directly between two points without occlusion, then we know that no opaque surfaces can be present on the line segment that connects those two points. Unfortunately reality is typically messier than this. Errors in visible surface estimation and shadow segmentation will invariably impact our ability to determine which voxels lie inside and outside of the true hidden scene.

In this section, we discuss how such errors manifest as reconstruction artifacts when a naive carving approach, such as the approach used in Section 4, is used. We then apply this insight to develop a probabilistic carving approach that is robust to pre-processing errors. We apply our robust method to reconstruct the shape of a mannequin using shadows cast onto non-planar relay surfaces that must be estimated from noisy depth measurements. We show that our robust algorithm handily outperforms a naive carving approach in this setting.

5.1 Sources of Error in Imaging Behind Occluders

False Carving: In a naive carving approach, a voxel is discarded from the set of points that lie inside a hidden object if just a single ray drawn from a laser spot to an observed “lit” pixel passes through that voxel. The reasoning behind this decision makes physical sense—if a voxel is occupied by an opaque material, then light should not be able to pass through it. In practice, this model is too strict. If the observed pixel is wrongly classified as lit when it in fact lies in shadow, then an entire ray of voxels might be erroneously carved out of the reconstructed scene. Furthermore, if there are errors in our estimates of the laser spot and observed pixel’s 3D locations, then the line segment that we carve through the scene will be misaligned with respect to the true propagation path. This will lead to further erroneous carving.

Complex light transport phenomena such as inter-reflections, specular reflections, translucent hidden objects or refraction can also cause false carving

because they cannot be explained by a two-bounce scattering model or opaque occlusions. An example of this is shown in Figure 6(a).

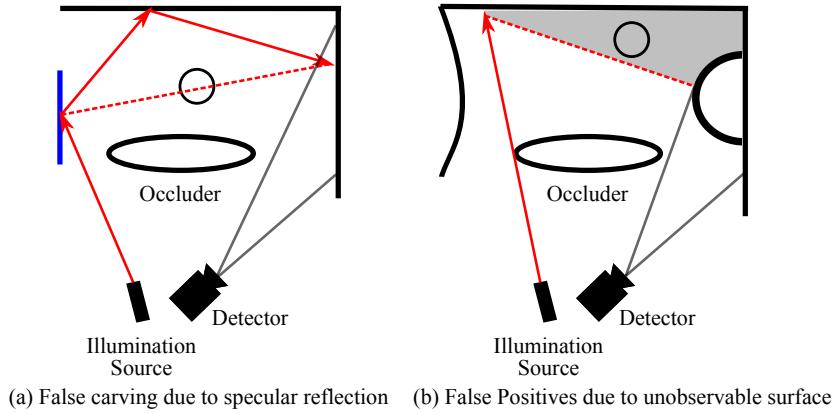


Fig. 6. Limitations on the hidden scene reconstruction. (a) Complex light transport such as specular reflection (blue surface) can result in unintended carving of the space occupied by the hidden object. (b) The shape of visible surfaces determines the recoverable space in the hidden volume.

False Positives: In addition to erroneously carving away voxels that are inside of a hidden object, a carving method may also fail to carve away a voxel that lies outside of all hidden objects. In some cases, these *false positives* cannot be prevented—for instance, if an empty voxel happens to lie inside of the hidden scene’s visual hull [13]. In other cases an empty voxel may remain uncarved for the same reasons that an occupied voxel may be carved erroneously; that is, due to errors in shadow classification or visible surface estimation. These sources of error tend to produce more false negatives (over-carving) than false positives when most voxels in the scene are probed by a large number of rays. This is because a naive carving method only requires a single *outside* result to carve away a voxel, whereas *all* tests must return an *inside* result for a voxel to be classified as outside of a hidden scene.

When the number of rays that probe a voxel is low, however, the probability of false positive classification becomes more significant. This becomes more of a problem when the visible surface geometry is complex, due to the fact that non-planar surfaces will often occlude themselves from the point of view of the observer. This self-occlusion reduces the set of points on the surface that can be illuminated and observed to probe the hidden scene. In many situations, this self-occlusion renders large swaths of the hidden scene unobservable. A simple example of this is shown in Figure 6(b). In general, however, any level of self-

occlusion will reduce the set of rays that can be probed and this will tend to increase the number of false positives in the reconstructed scene.

5.2 Robust Carving

To improve the robustness of our method to the false *outside* results that cause over-carving, we can relax our acceptance criteria by requiring some number $M > 1$ *outside* results to accrue before a voxel can be carved out of the hidden scene. This relaxed threshold reduces the probability of false carving at the cost of an increased false positive rate. This trade-off is often worth it if a stricter acceptance threshold would lead to excessive overcarving. To reduce the number of false positives, we can add a new acceptance criteria that requires some number $N > 0$ of *inside* results to accept a voxel as lying inside of the hidden scene. This acceptance criteria is motivated by the fact that an occupied voxel that *is* probed should produce *inside* results most of the time, in spite of any errors. This test also provides the added benefit of screening out all voxels in the hidden scene that are unobservable due to visible surface self-occlusion, or for other reasons.

Ideally, we would prefer to set these acceptance criteria using a principled approach that considers the probability of occupancy of each voxel. To this end, we considered a simple probabilistic model inspired by an acceptance test that was employed by Cheung et al. [7] for shape-from-silhouettes. If a voxel is subjected to N inside/outside tests, the result y_j of any single test is treated as a Bernoulli trial that is independent of the other $N - 1$ tests when conditioned on the true state of the probed voxel, which can be either empty (e) or occupied (o). Under this model the probability that a voxel is occupied given m outside results and n inside results is

$$\mathbb{P}(v_i = o | y_1, \dots, y_N) = \frac{\eta^m (1 - \eta)^n p_o}{(1 - \xi)^m \xi^n p_e + \eta^m (1 - \eta)^n p_o}. \quad (1)$$

Here η is the *miss probability*—that is, the probability that an occupied voxel projects to an illuminated region (is declared *outside*)—and ξ is the *probability of false alarm*—the probability that an empty voxel projects to a shadowed region (is declared *inside*). The values $p_o = \mathbb{P}(v_i = o)$ and $p_e = \mathbb{P}(v_i = e)$ represent the prior probabilities that voxel i is occupied or empty, respectively. A derivation of Eq. (1) is provided in the supplemental material.

To use Eq. (1) as an acceptance criterion, we can assume that all voxels share the same values of η , ξ , p_o , and p_e . Then we can calculate the conditional occupancy probability for all combinations of m and n that we might feasibly encounter in N observations (at most $(N+1)^2$ values) and store these probabilities in a lookup table. When all inside/outside tests have been completed, we look up the appropriate occupancy probability based on number of inside and outside counts, and accept a voxel as occupied if the probability lies above some threshold.

In theory, the parameters η , ξ , p_o , and p_e could be determined via an empirical analysis of inside/outside test errors. In practice we use them as tuning

parameters. A lower value of η signifies a higher level of trust in *outside* results. Reducing this parameter should reduce the rate of false positives and increase the rate of false carving. Setting η equal to zero reproduces the naive carving acceptance criteria. A lower value of ξ strengthens the effect of *inside* results. This means that fewer inside measurements will be required to declare a voxel occupied. The prior probabilities guide the acceptance criterion in the absence of measurement data, and in practice have very little effect on the acceptance result if a voxel has been probed more than a handful of times.

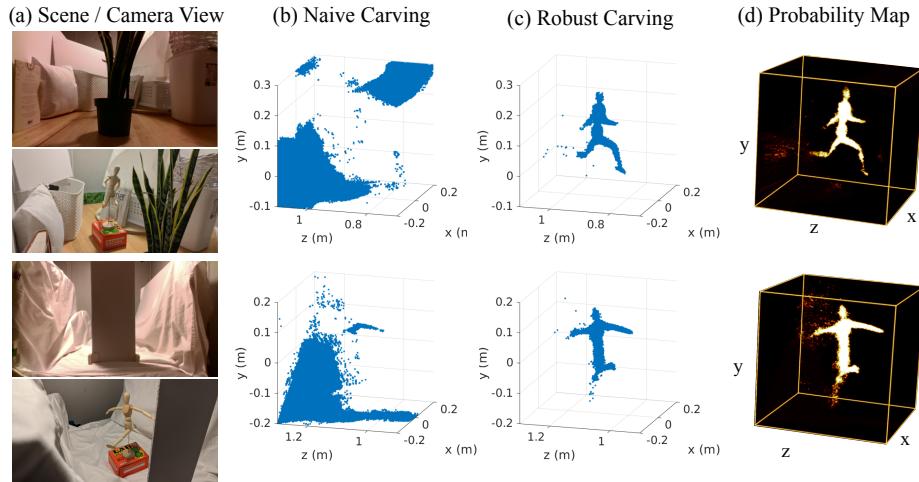


Fig. 7. Imaging behind occluders with complex visible geometries. (a) Scene setup and view from the camera. A mannequin is hidden behind an occluder. (b) A naive carving algorithm removes voxels occupied by the mannequin. (c) A robust carving method demonstrates robustness in reconstructing the hidden object with complex visible surfaces. (d) The probability map of voxel occupancy computed using Eq. (1).

5.3 Implementation

To test our robust carving method, we attempt to reconstruct the shape of a mannequin hidden behind an occluder by observing shadows cast onto non-planar and non-continuous visible surfaces consisting of various white objects (see supplemental material for results obtained when relay surface albedo varies spatially). We acquire the laser spot positions and the shape of the visible surface using a RealSense D435 active stereo RGB-D camera. We acquire shadow images and laser spot positions while the mannequin is present in the hidden scene. We then remove the mannequin to acquire a set of background images to aid in background segmentation. We perform shadow segmentation by applying a per-pixel threshold on the ratio of pixel values in shadow frames and background

frames. We note that we were also able to obtain good results without the aid of background measurements. These results as well as the shadow segmentation method that we used can be found in the supplemental material.

We present two reconstructions of the hidden mannequin. Each utilizes the same set of 30 illumination spots and shadow images. The first result, which is shown in 7(b), was produced by applying a naive carving criterion in which all voxels that are projected to an illuminated region of the visible surface *at least once* are carved away, and inside tests are not used. In this case, the mannequin reconstruction is clearly overcarved and resembles 3D pepper noise. The absence of inside tests also results in a residue of false positives at the periphery of the voxel grid. In Figure 7(c) we show results obtained when we applied the probabilistic acceptance criteria based on Eq. (1), and in Figure 7(d) we visualize the voxel occupancy probability map using a maximum intensity projection. In these results we can clearly distinguish the form of the mannequin.

6 Applications

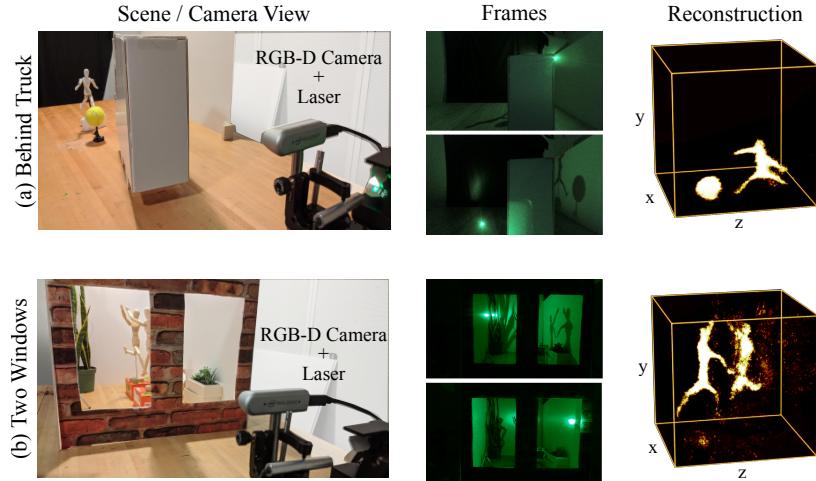


Fig. 8. Demonstration of potential applications of imaging behind occluders. Our technique can be used to (a) detect a person that is behind a truck and (b) reconstruct the a hidden part of rooms that have at least two windows.

We highlight some potential applications of our method by reconstructing hidden scenes embedded in settings that simulate situations that might be encountered in the real world.

Seeing Behind Trucks for Autonomous Vehicles: We want to simulate the ability to detect a child that is playing with a ball behind a parked truck and

is about to run out into the street. We observe a truck parked next to a building’s wall. A mannequin and ball were placed in front of the truck and out of view of the observer. We illuminate points distributed on the ground plane and on the building’s facade, and also observe shadows cast onto these surfaces. We acquire these surfaces with an RGB-D camera and use our robust carving method to reconstruct the shape of the reckless child. Results are shown in Figure 8(a).

Imaging Between Windows for Search and Rescue: We want to simulate a rescue team’s ability to locate humans in rooms. We observe a room with two windows from the outside. The room geometry has three vertical walls that are not parallel. The room also contains clutter seen in the left window. We shine the laser through one window and observe shadows cast onto a wall through the other window. As in the behind-trucks scene, we use an RGB-D camera to acquire the intensity and position of observed and illuminated pixels, and then use our robust carving method to reconstruct a hidden scene consisting of two humans in conflict. Results are shown in Figure 8(b).

7 Conclusion

We have introduced the novel NLOS imaging problem of *imaging behind an occluder*, and have demonstrated a method that solves this problem using a space carving approach that exploits observations of two-bounce light. We have modified our algorithm to be robust to errors in shadow segmentation and relay surface estimation, and this has enabled us to reconstruct hidden scenes embedded in reasonably complex visible scenes. As far as we are aware, we are the first to recover the detailed 3D shape of hidden, non-line-of-sight objects without the use of complex, specialized devices or significant calibrations, and are also the first to do so in real time and in non-trivial settings.

Our method can spur development in application areas such as new medical endoscopes, drones that can see behind trees while flying through forests, collision avoidance for autonomous vehicles, the rescue of humans hidden in rooms with more than one window, aerial imaging, industrial inspection, and more.

We hope to inspire more research in the imaging and computer vision communities that exploits the information conveyed by cast shadows. We also believe that our work highlights non-line-of-sight imaging as a new area of relevance for shape-from-silhouettes research.

In future work we hope to apply our method outside of the laboratory and in natural, larger-scale scenes. We believe that this is feasible if interference from ambient background light can be reduced by operating at infrared wavelengths or by employing time-modulated illumination. An exploration of how relay surface surface geometry affects reconstruction quality would also be valuable. Finally, variations on the *behind-occluders* problem could be explored such as reconstruction without depth information, or reconstructing transparent hidden objects.

Acknowledgements: We thank our reviewers for their helpful comments. This work was supported by DARPA REVEAL (N00014-18-1-2894) and the Media Lab Consortium. TS was supported in part by NSF GRFP (No. 1122374).

References

1. Ahn, B., Dave, A., Veeraraghavan, A., Gkioulekas, I., Sankaranarayanan, A.C.: Convolutional approximations to the general non-line-of-sight imaging operator. In: The IEEE International Conference on Computer Vision (ICCV) (2019)
2. Baradad, M., Ye, V., Yedidia, A.B., Durand, F., Freeman, W.T., Wornell, G.W., Torralba, A.: Inferring light fields from shadows. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
3. Batarseh, M., Sukhov, S., Shen, Z., Gemar, H., Rezvani, R., Dogariu, A.: Passive sensing around the corner using spatial coherence. *Nature communications* **9**(1), 1–6 (2018)
4. Baumgart, B.: Geometric Modeling for Computer Vision. Ph.D. thesis, Stanford University (1974)
5. Bouman, K.L., Ye, V., Yedidia, A.B., Durand, F., Wornell, G.W., Torralba, A., Freeman, W.T.: Turning corners into cameras: Principles and methods. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2270–2278 (2017)
6. Chen, W., Daneau, S., Mannan, F., Heide, F.: Steady-state non-line-of-sight imaging. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
7. Cheung, G.K., Kanade, T., Bouguet, J.Y., Holler, M.: A real time system for robust 3d voxel reconstruction of human motions. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR. vol. 2, pp. 714–720. IEEE (2000)
8. Franco, J.S., Boyer, E.: Fusion of multiview silhouette cues using a space occupancy grid. In: The IEEE International Conference on Computer Vision (ICCV). vol. 1. IEEE (2005)
9. Heide, F., O'Toole, M., Zang, K., Lindell, D.B., Diamond, S., Wetzstein, G.: Non-line-of-sight imaging with partial occluders and surface normals. *ACM Trans. Graph.* (2019)
10. Katz, O., Heidmann, P., Fink, M., Gigan, S.: Non-invasive real-time imaging through scattering layers and around corners via speckle correlations. *Nature Photonics* **8** (2014)
11. Klein, J., Peters, C., Martín, J., Laurenzis, M., Hullin, M.B.: Tracking objects outside the line of sight using 2d intensity images. *Scientific reports* **6**(1), 1–9 (2016)
12. Landabaso, J.L., Pardas, M., Casas, J.R.: Shape from inconsistent silhouette. *Computer Vision and Image Understanding* **112**(2), 210–224 (2008)
13. Laurentini, A.: How far 3d shapes can be understood from 2d silhouettes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **3**(2), 188–195 (1995)
14. Lazebnik, S., Furukawa, Y., Ponce, J.: Projective visual hulls. *Int. J. Comput. Vision* **74**(2), 137–165 (2007)
15. Liu, X., Guillén, I., La Manna, M., Nam, J.H., Reza, S.A., Le, T.H., Jarabo, A., Gutierrez, D., Velten, A.: Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature* pp. 1–4 (2019)
16. Maeda, T., Satat, G., Swedish, T., Sinha, L., Raskar, R.: Recent advances in imaging around corners. *arXiv preprint arXiv:1910.05613* (2019)
17. Martin, W., Aggarwal, J.K.: Volumetric descriptions of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **5**(2), 150–158 (1983)

18. Matusik, W., Buehler, C., McMillan, L., Gortler, S.: An efficient visual hull computation algorithm (2002)
19. Otsu, N.: A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics **9**(1), 62–66 (1979)
20. Raskar, R., Davis, J.: 5 d time-light transport matrix : What can we reason about scene properties? (2008)
21. Saunders, C., Murray-Bruce, J., Goyal, V.: Computational periscopy with an ordinary digital camera. Nature **565**, 472 (01 2019)
22. Savarese, S., Andreetto, M., Rushmeier, H., Bernardini, F., Perona, P.: 3d reconstruction by shadow carving: Theory and practical evaluation. International Journal of Computer Vision **71**(3), 305–336 (2007)
23. Tabb, A.: Shape from silhouette probability maps: Reconstruction of thin objects in the presence of silhouette extraction and calibration error. In: The IEEE International Conference on Computer Vision (ICCV). pp. 161–168 (2013)
24. Tancik, M., Satat, G., Raskar, R.: Flash photography for data-driven hidden scene recovery (2018)
25. Tarini, M., Callieri, M., Montani, C., Rocchini, C., Olsson, K., Persson, T.: Marching intersections: An efficient approach to shape-from-silhouette. In: In Proceedings of VMV 2002. pp. 255–262 (2002)
26. Thrampoulidis, C., Shulkind, G., Xu, F., Freeman, W.T., Shapiro, J.H., Torralba, A., Wong, F.N.C., Wornell, G.W.: Exploiting occlusion in non-line-of-sight active imaging. IEEE Transactions on Computational Imaging **4**, 419–431 (2018)
27. Velten, A., Wilwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M., Raskar, R.: Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. Nature Communications **3**(745) (2012)
28. Xin, S., Nousias, S., Kutulakos, K.N., Sankaranarayanan, A.C., Narasimhan, S.G., Gkioulekas, I.: A theory of fermat paths for non-line-of-sight shape reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6800–6809 (2019)
29. Yamazaki, S., Narasimhan, S.G., Baker, S., Kanade, T.: The theory and practice of coplanar shadowgram imaging for acquiring visual hulls of intricate objects. Int. J. Comput. Vision **81**(3), 259–280 (2009)