

# Reconstructing Scenes with Mirror and Glass Surfaces

THOMAS WHELAN, Facebook Reality Labs

MICHAEL GOESELE, Facebook Reality Labs and TU Darmstadt\*

STEVEN J. LOVEGROVE, JULIAN STRAUB, and SIMON GREEN, Facebook Reality Labs

RICHARD SZELISKI, Facebook

STEVEN BUTTERFIELD, SHOBHIT VERMA, and RICHARD NEWCOMBE, Facebook Reality Labs

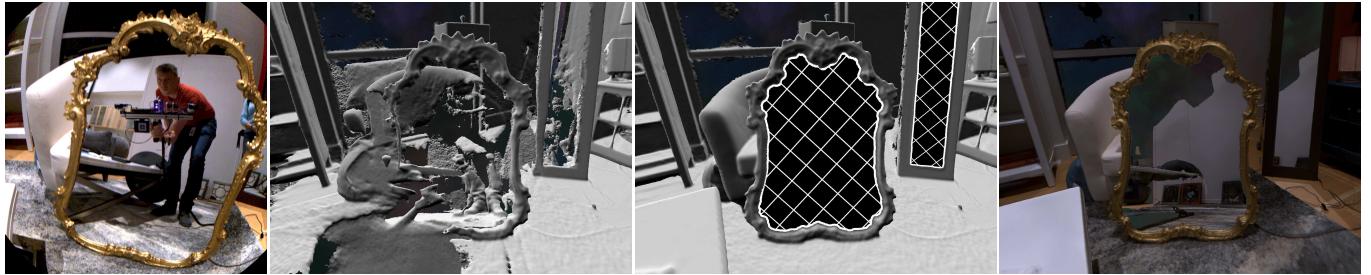


Fig. 1. Reconstructing a scene with mirrors. *From left to right:* Input color image showing the scanner with attached AprilTag in a mirror, reconstructed geometry without taking the mirrors into account, reconstruction taking the detected mirrors (rendered as cross-hatched area) into account and a photorealistic rendering of the scene including the mirrors. Detecting the mirrors is crucial for accurate geometry reconstruction and realistic rendering.

Planar reflective surfaces such as glass and mirrors are notoriously hard to reconstruct for most current 3D scanning techniques. When treated naïvely, they introduce duplicate scene structures, effectively destroying the reconstruction altogether. Our key insight is that an easy to identify structure attached to the scanner—in our case an AprilTag—can yield reliable information about the existence and the geometry of glass and mirror surfaces in a scene. We introduce a fully automatic pipeline that allows us to reconstruct the geometry and extent of planar glass and mirror surfaces while being able to distinguish between the two. Furthermore, our system can automatically segment observations of multiple reflective surfaces in a scene based on their estimated planes and locations. In the proposed setup, minimal additional hardware is needed to create high-quality results. We demonstrate this using reconstructions of several scenes with a variety of real mirrors and glass.

CCS Concepts: • Computing methodologies → 3D imaging; Shape modeling; Reflectance modeling;

Additional Key Words and Phrases: 3D scanning, reflective surfaces, mirrors, glass

## ACM Reference Format:

Thomas Whelan, Michael Goesele, Steven J. Lovegrove, Julian Straub, Simon Green, Richard Szeliski, Steven Butterfield, Shobhit Verma, and Richard Newcombe. 2018. Reconstructing Scenes with Mirror and Glass Surfaces. *ACM Trans. Graph.* 37, 4, Article 102 (August 2018), 11 pages. <https://doi.org/10.1145/3197517.3201319>

\*This work was carried out at Facebook Reality Labs.

Authors' addresses: Thomas Whelan, Facebook Reality Labs, [twhelan@fb.com](mailto:twhelan@fb.com); Michael Goesele, Facebook Reality Labs and TU Darmstadt; Steven J. Lovegrove; Julian Straub; Simon Green, Facebook Reality Labs; Richard Szeliski, Facebook; Steven Butterfield; Shobhit Verma; Richard Newcombe, Facebook Reality Labs.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *ACM Transactions on Graphics*, <https://doi.org/10.1145/3197517.3201319>.

## 1 INTRODUCTION

Active scanning techniques using Kinect-like sensors have recently been used very successfully to reconstruct indoor scenes [Kähler et al. 2015; Newcombe et al. 2011; Nießner et al. 2013]. One frequently occurring element of such scenes—reflective surfaces such as mirrors and glass—has, rarely been treated in previous work. While this at first sounds like a minor omission, mirrors actually pose a significant problem for any reconstruction system. A perfect mirror shows a perfect reflection of the world, which is indistinguishable from a direct observation of the mirrored world. The mirror is therefore essentially “invisible”. The mirrored scene, however, will still be reconstructed using standard vision techniques. Fig. 2 shows examples of bathrooms from the ScanNet dataset [Dai et al. 2017] with reflected scene parts reconstructed behind mirrors.

This mirrored geometry overlaps with real geometry located behind the mirror and interferes with its reconstruction. It is therefore not only desirable to reconstruct mirrors as standard elements in a scene, but also necessary in order to avoid invalid reconstructions of the overall scene. Similarly, glass surfaces are typically not reconstructed by sensors but should still be included in a reconstructed model. Due to the prevalence of mirrors and glass in common indoor environments, recent scene reconstruction approaches such as Matterport3D [Chang et al. 2017] require the user to manually select windows and mirrors in a scan.

In this paper, we propose a method to automatically detect and reconstruct mirrors and glass surfaces in a scene (see Fig. 1). Our key idea is to add a tag to the capture rig that can only be observed when the camera faces a mirror or glass surface. In our work, we use a mirrored version of an AprilTag [Olson 2011; Wang and Olson 2016]. Based on observations of this tag, we not only detect reflective surfaces, but also robustly estimate the plane parameters of the



Fig. 2. Example scans from the ScanNet dataset [Dai et al. 2017] with artefacts due to mirrors.

surface. Additionally, we develop a multiple-feature-based approach to detect the boundary of the planar mirroring surface.

Our contributions include:

- An automatic AprilTag-based approach, where we use a mirrored AprilTag to detect the existence of planar mirror or glass surfaces in the scene.
- An automatic bundle adjustment-based calibration approach to accurately detect the plane of the reflective surface as well as the relative tag location.
- An automatic approach to accurately reconstruct the boundary of framed and unframed mirror and glass surfaces and to distinguish between the two surface types.
- A thorough evaluation of our approach using quantitative measures as well as high quality renderings.

## 2 RELATED WORK

In this section, we survey related work on detecting and reconstructing reflective surfaces and objects, focusing on the work most relevant to our paper. We refer the reader to a recent state of the art report for more details on recovering the geometry of specular objects [Ihrke et al. 2010]. Our work uses a prototype 3D scanner, which implements a standard projector-camera system with an infrared pattern and truncated signed distance function (TSDF) based fusion. Since no modification to the scanning and reconstruction system is necessary, we refer the reader to the literature on KinectFusion and follow-up publications for details and related work [Kähler et al. 2015; Newcombe et al. 2011; Nießner et al. 2013].

### 2.1 Detecting Planar Reflective Surfaces

Planar glass and mirrors are difficult to detect in regular images. When light hits a glass surface, most of it enters the glass, is refracted, and leaves at a different location on the back. Only a small fraction of the light (typically less than 10%, as determined by the Fresnel equations) is reflected on the front and back surfaces and can be observed as two distinct, faint mirror images, unless it is masked by the scene behind the glass [Shih et al. 2015]. The fact that glass is a partially transparent and partially reflective surface provides valuable information since one can actually observe two distinct signals from the transmitted and reflected parts of the scene. This is widely used in multi-signal time of flight sensors [Foster et al. 2013; Jiang et al. 2017; Koch et al. 2017b,a]. Several passive imaging

approaches also rely on detecting and disentangling these two distinct image components. Applications include reflection removal [Arvanitopoulos et al. 2017; Xue et al. 2015] and scene reconstruction [Sinha et al. 2012; Wang et al. 2015; Wanner and Goldluecke 2013]. Other approaches rely on indirect observations, such as the fact that glass absorbs some light, especially in the infrared part of the spectrum [Klank et al. 2011] or changes the polarization state of the light [Miyazaki et al. 2004].

A standard mirror is covered by a thin reflecting layer at the back side of the glass. The reflection at this back side is typically much stronger than the front reflection, leading to the well known mirroring effect with hardly any observable double images. High quality mirrors for scientific applications have a reflective layer on the front surface (so called first surface mirrors), yielding an even crisper reflection. Given that we cannot observe two separate scenes, detecting mirror surfaces is very challenging. One option is to switch modalities, e.g., to use ultrasonic distance measurements, which are reflected from the mirror surface and can be reliably observed under specific imaging geometries [Yang and Wang 2008; Zhang et al. 2017].

In the optical domain, there are weak signals that can be used. For example, time of flight scanning systems that return multiple observations along the line of sight may return a weak signal reflected at the actual mirror surface. The presence of a framed mirror can also be inferred from the observed depth discontinuity (also called a jump edge) along the mirror-frame boundary [Käshammer and Nüchter 2015; Yang and Wang 2011].

Active and purely passive optical approaches can detect the mirror symmetry in captured images, depth images, or complete scene models [Yang and Wang 2011]. While this approach can in principle give a reliable indicator of the presence of a mirror in a scene, it fails if a part of the scene is only observed in the mirror and never seen directly. Finally, curved specular surfaces can be detected using their distinct distortion patterns [DelPozo and Savarese 2007].

### 2.2 Reconstructing Reflective Surfaces

The geometry of specular surfaces is typically recovered by detecting the specular reflection of an extended target, enabling the estimation of a normal direction to the surface observed at a given pixel. The key challenge of this approach is that the target needs to cover all relevant angles, which would in general require it to completely surround the object [Balzer et al. 2014]. Most practical approaches therefore use a limited-size target (see, e.g., [Liu et al. 2015; Tarini et al. 2005]) that can be moved around the object and recorded in multiple capture sessions [Balzer et al. 2011]. In the limit, one can even use the specular reflection of a single point light source aggregated over many frames [Chen et al. 2006]. Alternatively, passive capture approaches make use of the full environment reflected in a planar [Sinha et al. 2012] or curved and more complex surface [Godard et al. 2015]. For the special case of near-planar surfaces, Ding and Yu [2008] interpret the reflection as observed under orthographic projection as a general linear camera and determine its parameters from the observation of a checkerboard target. Jacquet et al. [2013] use reflections of lines in large window panes to reconstruct a normal field using a cut through the 3D video cube.

### 2.3 Scenes with Mirror and Glass Surfaces

For active scanning applications, most systems are able to scan through glass with a potential offset due to refraction. Scanning through a mirror surface is always possible as long as all light rays are reflected along the same planar mirror surface. This results in the reconstruction of a mirrored copy of the true scene geometry, located behind the mirror. Fasano et al. [2003] use mirrors in order to scan hard to reach areas using a laser stripe pattern scanner. Since regular mirrors provided low quality scan data, the authors use first surface mirrors. They also propose an option for a hand-held mirror, whose position they detect using markers on the mirror.

Foster et al. [2013] rely on the faint diffuse reflection observed on glass (and mirror) surfaces using a multi-return time of flight-based system. Given the physical reflection properties, glass can only be observed in this way under a very restricted set of angles. They extend the popular occupancy grid mapping, giving it the ability to include surfaces that are only observable in a fraction of the frames in which they should have been seen.

Jiang et al. [2017] use the intensity of the return signal for a time-of-flight (TOF) sensor, the distance to the surface, and the incident angle as features for a neural network-based classification of glass surfaces in indoor environments. While sufficient for robot navigation, their approach cannot yield high accuracy scan data.

Käshammer and Nüchter [2015] detect mirrors of known geometry in TOF data. They find boundaries of framed mirrors using the associated jump edges in depth. A similar metric is introduced by Yang and Wang [2011] to detect candidate mirror locations. We generalize the jump edge metric to also include frameless mirrors. Further, we use the reflection of an active tag in order to determine mirror planes. As a result, we are able to detect planar mirrors with unknown, arbitrary geometry in a scene.

An alternative line of work combines ultrasonic sensors with classical optical scanning in order to detect glass and mirror surfaces. Yang and Wang [2008] integrate a sonar sensor into a laser scanning system to detect glass. Evidence of mirrors is detected at depth discontinuities in the laser scans. The actual extent of a mirror is detected by performing an ICP based registration of real and reflected geometry.

Zhang et al. [2017] augment a Kinect scanning device with an ultrasonic distance sensor. They look for differences in optical and ultrasonic depths and use an elaborate MRF-based inference in order to detect the extent of glass and mirror surfaces. They can also reconstruct curved surfaces by fitting a parametric surface to the points from the acoustic sensor. One of the key limitations of ultrasonic depth sensors is that the surface needs to be observed at near-orthogonal angles, whereas our camera has a wide field of view and our tag is visible from a substantially larger set of orientations of the scanning rig.

## 3 OVERVIEW

The easiest way to identify a mirror or other semi-reflective surface such as glass in a scanning context is to identify the reflection of the scanning device itself in the mirror. While this is a classical vision problem with a certain expected error rate, detection can be simplified by adding easy to detect features such as a colorful

area to the scanner. Since the scanner itself is typically not part of the scene and needs to be removed from the scan (when directly observed or seen in a mirror), this will have no negative impact on the final scanned model.

We want, however, not only to detect the fact that a mirror is present but also to recover its position, orientation, and extent. For planar mirrors, this corresponds to finding the corresponding symmetry transform (to determine position and orientation) and defining the extent of the mirror plane. One way to solve the former problem is to attach a rigid calibration target to the scanning system that can be detected reliably and allows the stable matching of points on the target with feature locations in the recorded images [Rodrigues et al. 2010]. A simple prototype for such a system could be a standard cell phone that displays a calibration structure, records images with the front-facing camera, and performs all the processing. Once calibrated, such a system could reconstruct the mirror plane in the camera coordinate system from a single image.

In Sec. 4, we describe how we use an AprilTag-based calibration target attached to our scanner in order to detect glass and mirror surfaces and to reconstruct the mirror plane. We currently use the RGB camera of our scanning system for target detection, but this could also be performed in the infrared domain, where our pattern-based depth camera is operating. Since our actively illuminated (backlit) target emits light diffusely at a luminance level similar to the scene, it can easily be detected under a wide range of viewing directions while having minimal influence on scene illumination or, in the infrared case, the directionally emitted high frequency pattern of the projector.

Sec. 5 discusses how we jointly estimate the position of the target with respect to the scanner and the mirror plane. We then describe how we distinguish observations of multiple mirrors in order to reconstruct each of them separately (Sec. 6). In Sec. 7, we detail our approach to accurately detect the extent of the mirror before generalizing our method to also detect glass surfaces such as windows (Sec. 8). We evaluate the full system quantitatively and qualitatively for a wide range of scenes and reflective surfaces (Sec. 9) and conclude with an outlook on future work.

## 4 DETECTING A MIRROR

To ensure a robust detection of our target, we use AprilTags [Olson 2011; Wang and Olson 2016], fiducial markers frequently used in the robotics community. Detection using the authors' free implementation<sup>1</sup> happens in multiple phases. After some basic image processing for local intensity normalization, candidate locations are detected as continuous bright regions containing a dark region. Next, the edges and corners of the dark square region are extracted and the numerical code corresponding to the tag is robustly decoded. Finally, the edge locations are refined. Note that in order to directly use the existing AprilTag library, we manufactured a mirrored version of the tag that can only be detected when it is observed in a mirror. The output of the AprilTag library is the set of detected AprilTags per frame. Each detection contains the ordered image-location of the four corners and the tag's center, the tag ID, as well as additional information about the quality of the detection.

<sup>1</sup><https://april.eecs.umich.edu/software/apriltag/>

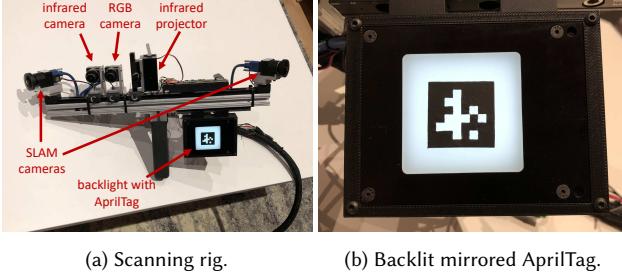
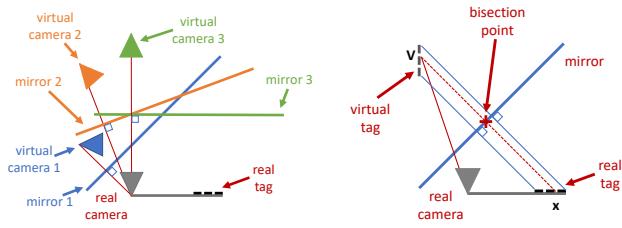


Fig. 3. The prototype of the scanning rig with attached backlit AprilTag.

Fig. 4. Initialization of the optimization. Left: Estimation of the mirror plane and the tag to camera transformation from at least three virtual camera estimations. Right: Estimation of the mirror plane from a single frame given known  $T_{ct}$ .

The AprilTags were constructed by applying 3M 468MP pressure sensitive adhesive to Thorlabs BKF12 black aluminum foil. The foil was then cut on a water jet and then sandwiched between sheets of ABS plastic with the adhesive side up. (Note that this supports only the manufacture of codes that consist of a single connected component.) The laminated pattern was then mounted on Metaphase backlights to ensure uniform illumination. We adjusted the intensity of the backlights to match our overall system setup, ensuring that no part of the captured tag is overexposed. Fig. 3 shows the mirrored AprilTag mounted onto the backlight and attached to our scanning rig. The width and height of the aluminum foil tag is 28.34 mm.

## 5 MIRROR SURFACE ESTIMATION

In the following, we assume that a typical SLAM system yields an accurate estimate of the pose of our scanning system for each frame [Engel et al. 2018; Mur-Artal et al. 2015]. We formulate the problem of estimating the mirror plane as minimizing the reprojection error of the corner points of the AprilTag mounted to the scan head into the camera view. This requires us to define the generative model for the projection, which we introduce in the following section.

### 5.1 Definitions

A transformation  $T_{ba}$  of a point from coordinate frame  $a$  into a coordinate frame  $b$  is represented as

$$T_{ba} = \begin{bmatrix} R_{ba} & a_b \\ 0^\top & 1 \end{bmatrix} \in \mathbb{SE}(3), \quad (1)$$

where  $R_{ba} \in \mathbb{SO}(3)$  and  $a_b \in \mathbb{R}^3$  is the origin of frame  $a$  expressed in frame  $b$  such that  $\dot{x}_b \propto T_{ba}\dot{x}_a$ . The points  $x_a$  and  $x_b$  are points in  $\mathbb{R}^3$ ;  $\dot{x}_a$  and  $\dot{x}_b$  represents their homogeneous lifting to  $\mathbb{R}^4$ . In our setting, we work with three coordinate frames:

- The corners of the AprilTag are defined in tag space  $t$  as  $c_t = (\pm\frac{\omega}{2}, \pm\frac{\omega}{2}, 0)$ .  $\omega$  denotes the width and height of the black portion of the AprilTag. The center of the AprilTag is located at the origin of  $t$ .
- The camera defines the camera frame  $c$  with the image plane aligned with the  $xy$  plane of the coordinate system.
- The world frame  $w$  is defined as the reference frame given by the SLAM system, which also yields the per-frame transformation  $T_{cw}^i$  from world to camera space as well as its inverse  $T_{wc}^i$ .

We define the plane of the mirror  $N$  using a standard plane parameterization as

$$N = \begin{bmatrix} \hat{n} \\ d \end{bmatrix} \in \mathbb{R}^4, \quad (2)$$

such that  $N \cdot \dot{x} = 0$  for points  $x$  on the plane.  $\hat{n}$  is the unit length normal of the plane and  $d$  is a scalar which represents the closest approach of the plane with the origin. Plane  $N_a$  in frame  $a$  can be expressed in frame  $b$  using the following expression:  $N_b = T_{ab}^\top N_a$ . Note the unusual use of the transformation matrix  $T_{ab}$  transposed and not obeying the index notation ordering.

The symmetry transform  $S$  that transforms a point  $\dot{x}_a$  into the virtual point  $\dot{v}_a$  as reflected in the planar mirror surface  $N_a$  via  $\dot{v}_a = S(N_a)\dot{x}_a$  is then given by

$$S(N) \equiv \begin{bmatrix} R(\hat{n}) & 2d\hat{n} \\ 0^\top & 1 \end{bmatrix} \in \mathbb{R}^{4 \times 4} \quad (3)$$

as described in Rodrigues et al. [2010]. The householder matrix

$$R(\hat{n}) \equiv \begin{bmatrix} I_{3 \times 3} - 2\hat{n}\hat{n}^\top \end{bmatrix} \in \mathbb{R}^{3 \times 3} \quad (4)$$

defines a reflection on a plane through the origin.

### 5.2 Generative Model

Given the above definitions, we can now write down the reprojection error for  $p_t^j$ , the  $j$ 'th tag corner (or tag center) in image  $i$ , using tag coordinates  $t$ :

$$f^{i,j}(\theta) = \pi_c(T_{cw}^i S(N_w) T_{wc}^i T_{ct} \dot{p}_t^j) - \tilde{p}_c^{i,j}, \quad (5)$$

where  $\tilde{p}_c$  is the observed (noisy) measurement of  $p_t^j$  in  $c$  and  $\pi_c$  is camera  $c$ 's projection function,  $\mathbb{R}^3 \rightarrow \mathbb{R}^2$ . We use the Kannala Brandt fish-eye projection model [Kannala and Brandt 2006] for this function.  $\theta$  is the optimization state vector consisting of the tag to camera transformation  $T_{ct}$  and the mirror plane  $N_w$ . Once the system is calibrated, we can treat  $T_{ct}$  as fixed and include only  $N_w$ . Combining all the observations of the tag yields the following sum-of-squares objective:

$$F(\theta) = \sum_{i=1}^N \sum_{j=1}^5 [f^{i,j}(\theta)]^2, \quad (6)$$

which we minimize using Ceres [Agarwal et al. 2018] with appropriate initialization as described in the following section. Updates to  $T_{ct}$  occur on the manifold of  $\mathbb{SE}(3)$  and we use a local minimal

parameterization for optimization,  $\Delta_{ct} \in \mathbb{se}(3)$  with an update step at each iteration of  $\mathbf{T}_{ct} \leftarrow \mathbf{T}_{ct} e^{\Delta_{ct}}$ . We leave the homogeneous representation  $\mathbf{N}_w$  over-parameterized and divide by the magnitude of the first three components after convergence to reach the canonical representation as shown in Eq. 2.

Although we do not do so here, our reprojection error could also be included as a term within a classic SLAM or structure from motion system to improve localization and provide more thorough Bayesian treatment of the complete acquisition system.

### 5.3 Initialization and Calibration

Since Eq. 6, which we seek to minimize, is non-convex and Ceres relies on gradient-based optimization, we must ensure an adequate initialization of the parameters to reach a globally optimal solution. Rodrigues et al. [2010] describe the calibration of a system with a fixed camera, a fixed object outside the camera’s view, and a moving planar mirror, through which the camera can observe the object. This is equivalent to our case of the scan system with a rigidly connected camera and tag moving in front of a static mirror.

Following their approach, given known intrinsic calibration parameters for camera  $c$  and known dimensions of the AprilTag, we compute the camera from virtual-target transform  $\mathbf{T}_{ctv}^i$  using standard exterior pose from known correspondence [Zhang 2000] for each image of the tag in frame  $i$ . This is depicted to the left in Fig. 4. The output of the exterior pose estimation is a 6 degrees-of-freedom transform, but the mirror plane minimally parameterized in  $\mathbb{R}^3$  has only 3 degrees-of-freedom. The subset of  $\mathbb{SE}(3)$  in which we observe motion informs us of the mirror plane. Given at least three such transforms  $\mathbf{T}_{ctv}^1, \mathbf{T}_{ctv}^2, \mathbf{T}_{ctv}^3$ , we can estimate the initial values for the mirror plane  $\tilde{\mathbf{N}}_w$  which in turn allows us to compute the real tag to camera transform  $\tilde{\mathbf{T}}_{ct}$ . In practice, we observed that it is sufficient to estimate  $\tilde{\mathbf{N}}_w$  and to set  $\tilde{\mathbf{T}}_{ct}$  to the identity to initialize our non-linear refinement.

### 5.4 Single Image Mirror Plane Estimation

Given the rigidity of the scanning system, we can calibrate the tag to camera transform  $\mathbf{T}_{ct}$  once by minimizing Eq. 6 and freezing the result. This allows us to estimate the mirror plane  $\tilde{\mathbf{N}}_w$  from a single frame. Similar to above, we estimate the pose  $\mathbf{T}_{ctv}$  of the virtual (i.e., reflected) AprilTag with respect to camera  $c$  given  $c$ ’s intrinsic parameters and the AprilTag’s known dimension  $\omega$ . The mirror plane can now be established as that which bisects the corresponding real and virtual corner locations as depicted in Fig. 4 to the right. We use these single image mirror estimations as input for the next section, where we perform a clustering step to split observations from multiple mirrors and to estimate the final plane parameters for each mirror.

Note that we only need to see the AprilTag in a single image and can then transfer the observation information to all other frames using the scene geometry and SLAM poses. Observing the mirror (and potentially the AprilTag) in additional frames may increase the accuracy of the estimated mirror plane and helps in accurately detecting the mirror boundary as described in Sec. 7.

## 6 MULTI-MIRROR PARAMETER ESTIMATION

We proceed by assuming that each AprilTag observation  $i$  belongs to its own mirror for which we can then estimate  $\mathbf{N}_w^i$  as discussed previously. Additionally, we can intersect the center of the 2D tag detection with the estimated plane to find the observation point on the surface of the mirror through which our AprilTag has been reflected. Given the low false-detection rate of AprilTags, we have high confidence that this point belongs to a mirror surface. For each observation, we now have a point and normal pair,  $[\mathbf{P}_w^i, \hat{\mathbf{n}}_w^i]$ .

To separate the observations into sets for each mirror in the scene, we use a non-parametric clustering algorithm that we denote as DP-planes, since it is derived from DP-means [Kulis and Jordan 2011] and DP-vMF-means [Straub et al. 2015]. Except for using different distance metrics, these algorithms perform the same k-means-like alternating optimization: In step (1): incrementally assign data points to the closest existing cluster unless the closest cluster is further than some threshold  $\lambda$ . In the latter case, a new cluster is instantiated from the query data point. In step (2): recompute the cluster centers given all associated data points. For a more detailed algorithm description, we refer to the original papers. We take the existing algorithm but modify the distance metric to a symmetrized point-to-plane distance:

$$\text{dist}([\mathbf{p}_a, \hat{\mathbf{n}}_a], [\mathbf{p}_b, \hat{\mathbf{n}}_b]) = \frac{1}{2} (\|(\mathbf{p}_a - \mathbf{p}_b) \cdot \hat{\mathbf{n}}_b\| + \|(\mathbf{p}_b - \mathbf{p}_a) \cdot \hat{\mathbf{n}}_a\|). \quad (7)$$

This distance measures how compatible two planar observations are to one another and does not require that we choose an arbitrary weighting between the angular and Euclidean components of the observations. In step (2) of the DP-planes algorithm cluster centers in the planar space are computed from the set of points  $p_i$  and normals  $n_i$  in cluster  $k$ ,  $I_k = \{p_i, n_i\}_{i=1}^{N_k}$ , as

$$\bar{\mathbf{p}}_k = \frac{\sum_{i \in I_k} \mathbf{p}_i}{N_k} \quad \bar{\mathbf{n}}_k = \frac{\sum_{i \in I_k} \mathbf{n}_i}{\|\sum_{i \in I_k} \mathbf{n}_i\|}. \quad (8)$$

As in the aforementioned algorithms, the DP-planes algorithm assigns data points to cluster centers via the symmetric point-to-plane distance if the distance is within some threshold  $\lambda$ , measured in meters. If the distance exceeds this threshold, a new cluster is initialized with the value of the observation. We find  $\lambda = 10\text{ cm}$  to yield good mirror-plane segmentation in all our experiments with non-coplanar mirrors. For coplanar mirrors, we can refine the segmentation using the boundary detection (see Sec. 7) and recompute the mirror plane parameters for newly segmented mirrors.

Equation 8 directly provides a joint estimate of the plane parameters for each individual mirror, derived from the previously independently estimated oriented surface samples. We use these for all subsequent steps.

## 7 MIRROR BOUNDARY DETECTION

Given the mirror plane estimated in Sec. 5, we use it as a natural parameterization for boundary extraction. We discretize the plane into a grid with square cells, typically using a resolution of  $5\text{ mm}$ , project all features discussed in the following onto this plane, and extract the boundary using a total variation-based segmentation.

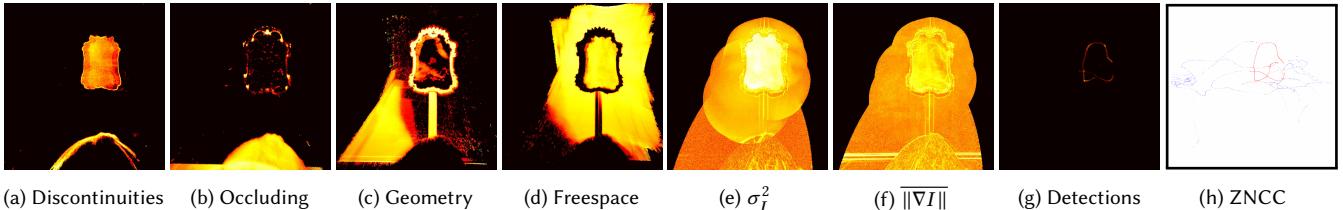


Fig. 5. We explore eight different feature channels to facilitate mirror segmentation. All channels are shown in log-scale using the “hot” colorscheme except the rightmost ZNCC channel which is displayed from  $-1$  in blue to  $1$  in red. The first four features are computed from the depth image. The discontinuities channel (a) indicates the mirror boundaries and Occluding (b), Geometry (c) and Freespace (b) indicate structures in front of, right around and behind the mirror plane. Feature channels (e) and (f) aim to extract mirror boundary information from image intensities: high intensity variance  $\sigma_I^2$  indicates a reflective surface and high average intensity gradient  $\|\nabla I\|$  is expected at the mirror boundary. Channel Detections (g) accumulates the AprilTag detections. Using the zero-mean normalized cross-correlation (ZNCC) of the AprilTag appearance with the actual predicted image intensities, the ZNCC channel (h) allows us to harvest non-mirror detections at low ZNCC-valued areas.

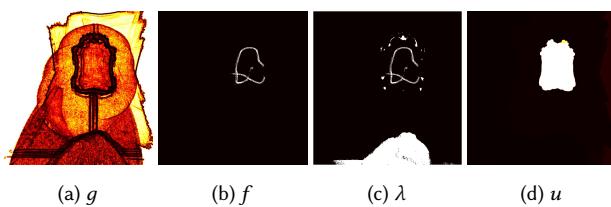


Fig. 6. Left to right: The boundary weighting term  $g(x)$ , the segmentation constraint image  $f$ , the constraint weighting image  $\lambda$ , and the resulting mirror segmentation image  $u$ .

## 7.1 Feature Extraction

In the discretized mirror plane, we compute three sets of feature channels. The first set is derived from geometry information, the second set is based on image intensities, and the third set depends on observations of the AprilTag. Fig. 5 visualizes the different channels in log-space for a baroque-style (ornate edge) mirror (see Fig. 8j).

**Geometric Features.** For the geometric features, we determine for each depth sample the intersection of the ray from the camera of the depth sensors with the mirror plane, and increment its intersection count. We then classify each depth sample according to its signed distance  $d$  to the mirror plane as **Occluding** (for  $d > \delta$ , i.e., for samples in front of the mirror plane), **Geometry** close to the mirror plane (for  $-\delta \leq d \leq \delta$ ) or **Freespace** further away than the mirror plane (for  $d < -\delta$ ). Positive distance values indicate a sample in front of the mirror plane. We use a threshold of  $\delta = 20\text{ mm}$  to capture depth and pose estimation noise as well as calibration inaccuracies. Each of the features is defined as the ratio of classified sample count to total intersection count for each cell.

One characteristic of mirrors is that they create depth discontinuities at their border between the reflected scene and the frame for framed mirrors [Käshammer and Nüchter 2015]. Frameless mirrors create a depth discontinuity between the reflected scene and the scene behind the mirror. We capture both by determining the ratio of **Discontinuities** in a cell, aggregated as above with one difference: Since a discontinuity appears in a depth map as soon as a boundary is seen from the camera or the projector, we additionally also determine for each sample the intersection of the projector sample ray

with the mirror plane and accumulate counts also for this cell. We define a depth sample as belonging to a depth discontinuity if the range of depth values in its  $9 \times 9$  sample neighborhood in the depth map exceeds  $10\text{ cm}$ . We use a fairly large neighborhood since depth samples right at the boundary are often not reconstructed.

**Intensity-based Features.** To further constrain the mirror segmentation, especially in the case of frameless mirrors, we consider two intensity-based feature channels that use the projection of the color images on the mirror plane: We compute the **Intensity Variance**  $\sigma_I^2$  for each cell, i.e., the variance of the intensities projected onto that cell. Because of the variability of the reflection in the mirror, we expect high variance inside the mirror. In addition, we observe higher variance for all non-reflected scene parts that are not in the mirror plane. This feature is thus related to the geometric, occluding and freespace features.

The **Mean Intensity Gradient**  $\|\nabla I\|$  corresponds to the geometric discontinuities channel. For each cell in the mirror plane, we average the image gradient norm. Since two different parts of the scene are observed at the boundary of the mirror, this leads to a high average intensity gradient.

**Observation-based Features.** The AprilTag itself yields valuable information. Given the properties of the AprilTag detector, we mark for each **Detection** the cells covering the locations of the four corners and the center of the Apriltag in the mirror plane. These provide strong positive evidence of a mirror.

Finally, we compute where in an image we would see the AprilTag if it would be reflected in a mirror. We compute the average zero-mean normalized cross-correlation **ZNCC** [Brown 1992; Lewis 1995] between the average tag appearance and the area in the current image at the predicted tag location assuming reflection about the mirror plane. This channel allows us in particular to harvest non-mirror areas as indicated by low ZNCC scores.

## 7.2 Boundary Extraction

Given the features described in the previous section, we perform  $g$ -weighted Total Variation segmentation [Unger et al. 2008a,b] (detailed below), which has been used successfully in semi-supervised image segmentation. A binary mirror/non-mirror segmentation is

relaxed to real values between 0 and 1. The segmentation  $u$  is defined as the solution to the following minimization over the mirror image space  $\Omega$ :

$$u^* = \arg \min_{u \in [0, 1]} \int_{\Omega} g(x) \|\nabla u(x)\| d\Omega + \int_{\Omega} \lambda(x) |u(x) - f(x)|_1 d\Omega. \quad (9)$$

$g$  weights the boundary length regularization in the first term to encourage the boundary to lead through low values in  $g$ . In the second term, the segmentation  $u$  is constrained to be close to the function  $f$  through the  $L_1$  norm weighted by  $\lambda$ . Higher values of  $\lambda$  lead to a stronger constraint on the segmentation. If  $\lambda$  is 0, constraints are not enforced.

We compute the boundary weighting term  $g$  from a combination of the feature channels  $c_i$  as

$$g(x) = \exp(-\sum_{i \in \mathcal{I}} \alpha_i \|\nabla \log(c_i(x))\|^{\gamma_i}), \quad (10)$$

where we use the set of channels  $\mathcal{I} = \{\text{Discontinuities, Geometry, Freespace}, \sigma_I^2, \|\nabla I\|\}$  and the tuned coefficients  $\alpha = \{0.04, 0.125, 0.05, 0.05, 0.1\}$  and  $\gamma_i = 0.8$ . This encapsulates the notion that we want the mirror boundary in areas where the gradients of the feature channels are high, as can be seen from Fig. 5 and the combined  $g$  in Fig. 6a.

Using the ZNCC channel and the AprilTag detections, we set  $f(x)$  and  $\lambda(x)$  at pixel location  $x$  to constrain the segmentation to be 1 at tag detections and 0 wherever occluding structure was detected and at the discretized mirror plane boundaries. Additionally, we incorporate weak mirror/non-mirror detections from the ZNCC feature channel:

- (1)  $f(x) = 1$  and  $\lambda(x) = 10^3$  for all target detections indicated in the detection feature channel,
- (2)  $f(x) = 1$  and  $\lambda(x) = 10^{-1}$  for cells with ZNCC value  $> 0.8$ ,
- (3)  $f(x) = 0$  and  $\lambda(x) = 10^{-1}$  for cells with ZNCC value  $< -0.2$ ,
- (4)  $f(x) = 0$  and  $\lambda(x) = 10^3$  for the boundary of the cell domain,
- (5)  $f(x) = 0$  and  $\lambda(x) = 10^3$  for cells with high Occlusion value,
- (6)  $f(x) = 0$  and  $\lambda(x) = 0$  for all other pixels.

We use aggressive thresholds and a small  $\lambda$  value for the ZNCC-derived constraints to reflect the lower confidence in them, since they are influenced by errors in the overall system. This ensures a low rate of misclassifications.

Equation 9 can be optimized efficiently and optimally as described in Unger et al. [2008a] using a primal-dual approach yielding the segmentations shown in Fig. 6. We use  $\theta = 0.1$  and  $\tau = 0.2$  as proposed by Unger et al. [2008a] and iterate our GPU-based implementation for 10,000 iterations to ensure convergence.

We use marching squares to extract a sub-cell accurate piece-wise linear mirror boundary as the iso-contour at value 0.5 in the segmentation image  $u$ . The marching squares algorithm is the equivalent of marching cubes [Lorensen and Cline 1987] on a 2D grid.

## 8 HANDLING GLASS

As discussed in the previous sections, glass surfaces differ from mirrors in multiple ways. First, images of a glass surface are in general a mixture between the transmitted and reflected scenes. The reflected image is therefore both diminished in brightness and potentially corrupted with the texture from the direct light path.



Fig. 7. Examples of offset observed on glass at varying distances.

Any feature detection in the reflected scene must therefore be robust to relatively low contrast and signal to noise ratio.

Second, the reflected scene is reflected on the front and back surfaces of the glass, yielding a double image. This effect depends on the distance of the scanning rig from the glass surface (see Appendix A and Fig. 7). In our experience, the AprilTag library will not detect tags if the offset is too large and will otherwise typically reconstruct one of the two tag locations. It is therefore sufficient to keep a minimum distance from the glass pane while scanning.

Third, we need to distinguish between glass and mirror surfaces. Our implementation classifies a surface as glass if we observe geometry within the projected area of the detected AprilTag that is neither at the depth of the AprilTag nor within  $\delta$  of the reflective plane. This is only possible for glass whereas for a mirror, the AprilTag serves as an occluder. In other words, detection of geometry through the image of the AprilTag implies we are seeing past the surface. We note that this distinction will fail for shallow objects such as picture frames leading to a misclassification of a glass surface as mirror, shown in Fig. 15d. An alternative classification approach would be to detect the intensity of the reflected AprilTag, which is significantly lower for glass than for a mirror.

Apart from these changes, our pipeline is directly able to reconstruct the plane as well as the boundary of framed glass surfaces as we will show in the following section.

## 9 RESULTS

We implemented our reconstruction pipeline on a 6 core Intel Core i7-5930K system with an NVIDIA TITAN Xp GPU and Ubuntu 16.04. The depth maps have a resolution of  $960 \times 640$  pixels; the RGB images have a resolution of  $1224 \times 1024$  pixels. The baseline reconstruction system (depth extraction, depth fusion, geometry extraction using dual contouring, texture generation but excluding SLAM) runs on this configuration at  $\approx 37$  Hz. Using 12 threads, we can estimate the AprilTag locations in the RGB images at  $\approx 70$  Hz. The feature computation for boundary extraction runs at  $\approx 38$  Hz. The throughput of the boundary segmentation optimization is  $\approx 60k$  pixels per second such that a  $640 \times 480$  pixel set of feature channels takes  $\approx 5.12$  s to segment. Overall, reconstructing a mirror area of  $\approx 0.5 m^2$  from 700 frames takes about 90 s. We used identical parameters for all results (mirrors as well as glass).

We demonstrate our system on a wide variety of mirrors and glass surfaces (see Fig. 8): a flat surface mirror (Fig. 8a), frameless mirrors (Fig. 8b–8d), a beveled mirror (Fig. 8e), framed mirrors (Fig. 8f–8m), a frame mirror with texture on the mirror surface (Fig. 8h), a slightly bent mirror (Fig. 8n), and glass surfaces (Fig. 8o–8r).

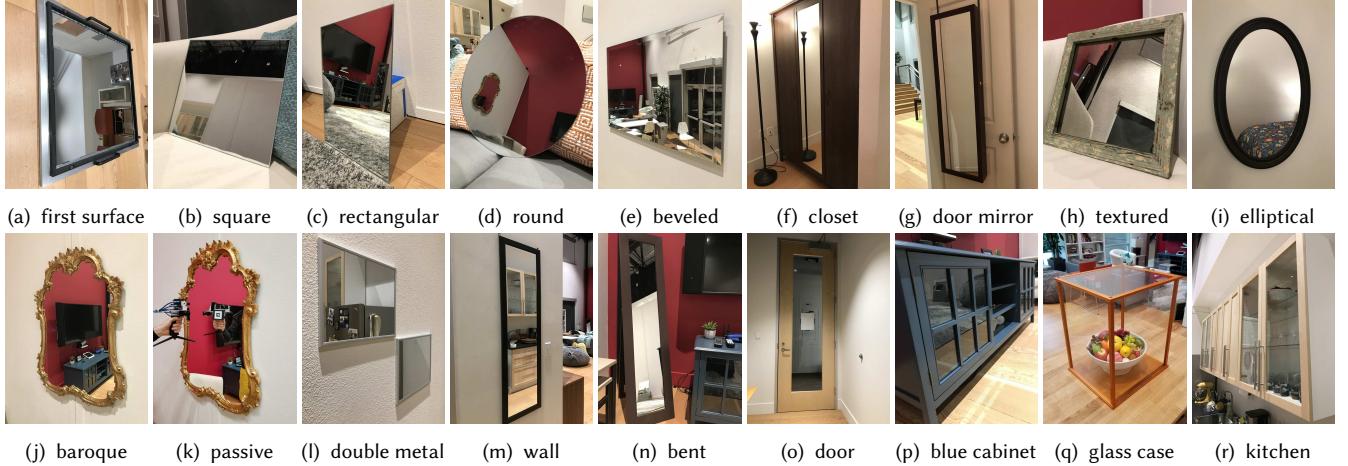


Fig. 8. Overview of the mirrors and glass surfaces used in our experiments. The first surface mirror 8a serves as ground truth flat mirror. Mirrors 8b–8e are frameless mirrors whereas 8e is frameless with a bevel. Mirrors 8f–8m are framed. The textured mirror 8h has a printed noise texture on the mirror surface, which is faintly visible in the image. The duplicate baroque mirror 8k is reconstructed with the illumination on the tag switched off, functioning as a passive non-emitting tag. The double metal mirrors 8l are low quality metal-only mirrors (i.e., not based on glass). Mirror 8n is slightly bent, yielding a slimming effect. Finally, 8o–8r show examples of glass surfaces captured.

Table 1. Error metrics (RMS error) for our reconstructions for a plane estimated from a single observation (Sec. 5.4), for a plane determined by the cluster center (Sec. 6) and for a plane estimated using the full optimization (Sec. 5.3). Reprojection errors are given in pixels; geometric errors in mm. All values are RMS errors. The geometric error for single observations is always zero up to numerical precision. For the datasets marked with \*, the errors are accumulated over all reflective surfaces.

dataset	single obs.		clustering		full estimation	
	reproj. (pixel)	reproj. (pixel)	reproj. (pixel)	geom. (mm)	reproj. (pixel)	geom. (mm)
first surface	0.066	0.34	2.50	0.21	2.66	
square	0.063	0.22	2.14	0.19	3.17	
rectangular	0.056	0.36	1.77	0.34	3.51	
round	0.061	0.55	1.72	0.51	2.23	
beveled	0.058	1.39	2.04	1.35	8.17	
closet	0.065	0.73	2.86	0.72	6.37	
door mirror	0.075	0.92	3.04	0.92	3.05	
textured	0.063	0.15	3.53	0.15	3.56	
elliptical	0.065	0.25	2.77	0.22	3.67	
baroque	0.059	0.54	2.11	0.50	5.64	
passive	0.066	0.92	2.76	0.56	4.67	
double metal*	0.067	1.39	3.40	1.42	3.91	
wall	0.078	3.17	1.43	2.84	16.74	
bent	0.075	5.61	3.54	6.33	29.46	
door	0.28	0.58	4.68	0.50	4.77	
blue cabinet*	0.085	1.41	2.92	1.27	5.98	
glass case*	0.383	9.56	8.3	3.58	8.74	
kitchen*	0.097	1.62	3.28	1.54	4.16	

## 9.1 Quantitative Results

In order to quantitatively evaluate our method, we evaluate the reprojection error and the geometric error after multiple stages of our pipeline. As shown in Table 1, we can accurately estimate the

mirror plane from a single observation for all datasets. Only the door and the glass case show a noticeably larger error. The error of the clustering-based estimation, which jointly estimates a single plane for all observations, increases significantly as expected. The joint full estimation based on reprojection error is able to minimize it while typically increasing the geometric error. This is especially pronounced for the wall and bent standing mirror and the datasets with multiple glass panes. We also note that the first surface mirror yields one of the lowest geometric errors after full estimation.

## 9.2 Reconstructions

In Fig. 9, we show a side by side comparison of a real world input image and the reconstruction produced by our method. Fig. 10 shows a full scene reconstruction rendered from a novel point of view. In Figs. 1 and 11 through 13, we show reconstructed surfaces containing mirrors or glass that produce erroneous geometry when not properly handling mirrors in the left column, the detected mirror surface and segmentation in the middle, and the rendered scene given the reconstructed mirror on the right.

For all mirror examples naïvely, fusing the depth images produces poor reconstructions with holes where there should have been surfaces and erroneous reflected geometry behind the mirror plane. While the segmentation of the frameless mirrors in Fig. 11 is not perfect around the boundaries, it still allows us to faithfully reconstruct the scene. Note that previous work is completely unable to handle such cases automatically. Our mirror segmentation can handle arbitrary shaped mirror boundaries as can be seen in the baroque-style mirror in Fig. 1. Interestingly, for the textured mirror in Fig. 12, the naive depth fusion actually partially reconstructs the mirror surfaces due to partial reflections of the IR dot pattern on the texture. However, as can be seen, this does not disturb the proposed



Fig. 9. Side by side rendering of the real world input image on the left and our reconstruction and rendering on the right.



Fig. 10. Full scene reconstruction showing multiple reconstructed mirrors and their interreflections.

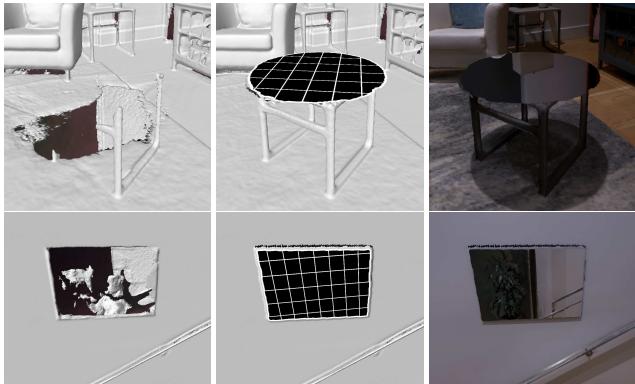


Fig. 11. Top row: frameless round mirror used as a table top (cf. Fig. 8d). Bottom row: beveled mirror mounted on a wall (cf. Fig. 8e). From left to right: Reconstructed geometry without taking mirrors into account, reconstruction taking the mirrors into account, and photorealistic rendering including the mirrors.

mirror segmentation pipeline. The slightly bent mirror in Fig. 12 is properly approximated as a planar mirror by the proposed system.

The glass cupboard windows in Fig. 13 are successfully reconstructed, segmented and classified as glass. Note that the pottery inside the cupboard is reconstructed accurately through the glass. Although complex in nature in terms of visible reflections, our system is able to reconstruct the glass museum display case with five glass panes shown in Fig. 13 without any modifications.

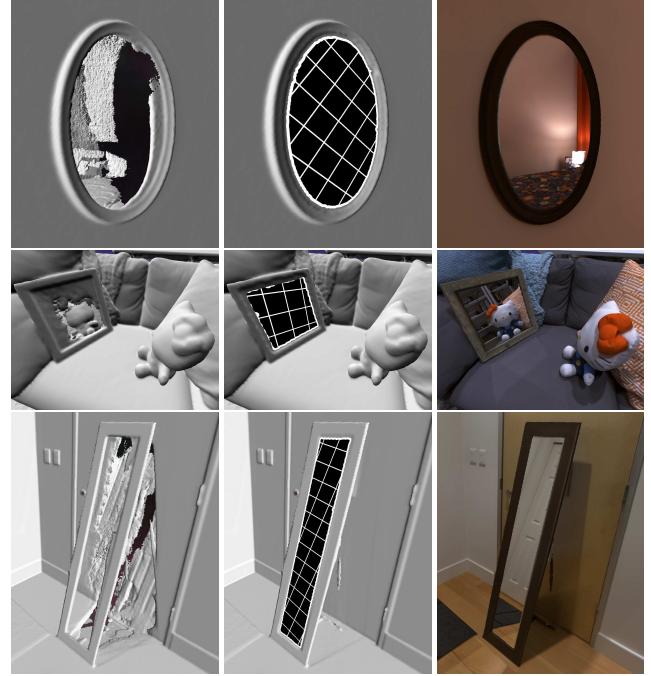


Fig. 12. Top row: Framed round mirror hanging on a wall (cf. Fig. 8i). Middle row: Framed mirror with some slight texture on its surface (cf. Fig. 8h). Bottom row: Slightly bent free standing mirror with a frame (cf. Fig. 8n).

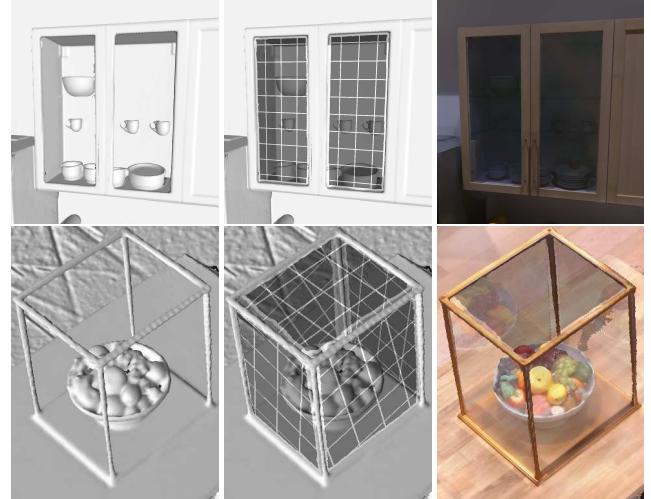


Fig. 13. Top row: Cupboard with glass windows (cf. Fig. 8r). Bottom row: Glass museum case (cf. Fig. 8q).

We show in Fig. 14 that our approach does not require a backlit tag to achieve accurate results. In this sequence, we rely only on the ambient light available in the scene to illuminate the target. This demonstrates that our technique also works with a simple matte printout of an AprilTag and does not depend on difficult to source custom hardware.

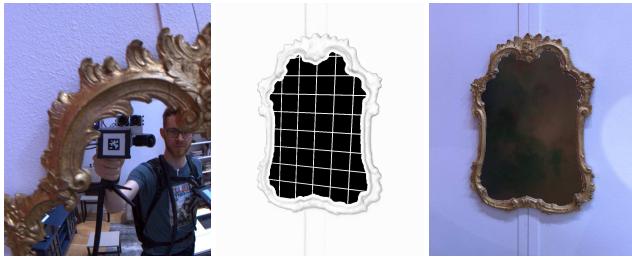


Fig. 14. Alternative reconstruction of the baroque mirror using a passive (i.e., non-illuminated) tag (see Figs. 1 and 8k). *From left to right:* Sample frame from the input image sequence, reconstructed geometry, and a textured reflecting reconstruction.

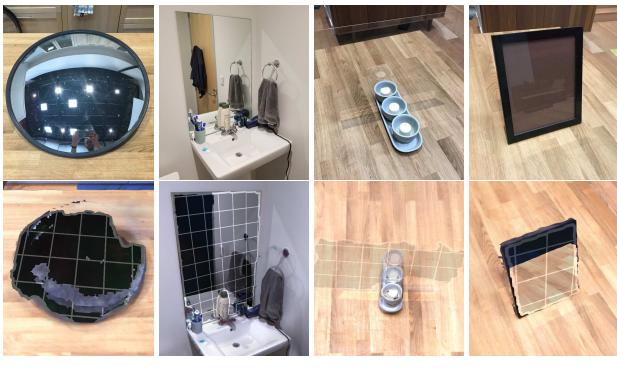


Fig. 15. Four examples of challenging structures that result in varying degrees of failure. The top row shows real photographs while the bottom row shows the output of our system. These failures fall into three categories: non-planar mirror geometry (a), lack of border observability (b, c), and incorrect glass-mirror classification (d).

### 9.3 Limitations and Failure Cases

While our approach is in general highly robust, we still observed occasional failure cases at various stages of the processing pipeline. If the AprilTag is not detected in any of the input frames, our approach fails catastrophically. This is typically caused by bad imaging conditions such as blurred images due to fast scanner movement, only partial visibility of the tag, low contrast (in particular on glass surfaces with both passive and back-lit targets, see also Fig. 7) or highly curved reflective surfaces. This could be alleviated by a more visible target, e.g., a set of LEDs marking corners of a planar tag.

Given an AprilTag detection, we can always reconstruct a mirror plane for a single observation using the approach described in Sec. 5.4. For slightly curved mirrors, approximate reconstruction is possible as our approach will often produce a plausible plane fit (e.g. Fig. 12, bottom). However, for a strongly curved mirror, our representation is unable to produce an accurate estimate of the surface, as shown in Fig. 15a (not visible is a phantom mirror plane that our approach also estimated to lie behind the surface due to clustering of highly distorted tag reflections). The model we use

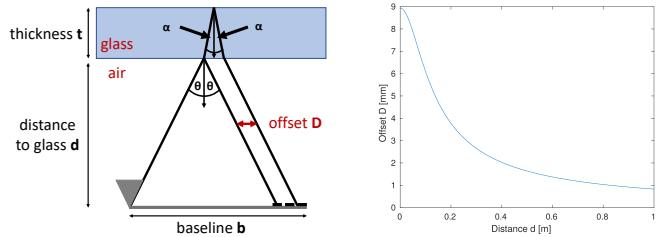


Fig. 16. Left: The glass configuration. Right: The offset in mm.

could be extended to account for this but would require a denser set of observations to resolve the surface shape.

Compared to the plane estimation, detecting the boundary is much more challenging since it relies on more subtle cues as can be seen in many examples in this paper. In situations where the border is occluded, for example in the bathroom scene in Fig. 15b, our approach will not try to infer hidden structure and only resolves the boundary up to the regularization capabilities of the segmentation. Borderless glass presents a challenging case where the photometric cues are too weak to constrain the boundary, shown in Fig. 15c.

As mentioned in Sec. 8, a failure case in our glass classification occurs when there is geometry within  $\delta$  of the estimated plane. This is demonstrated in Fig. 15d with a picture frame glass that is classified as mirror. As discussed previously, a remedy to this problem would be to calibrate the reflected intensity of the tag on mirrors and glass and use that cue to distinguish between the two, as a reflection from glass would be significantly darker than one from a mirror.

## 10 CONCLUSION AND FUTURE WORK

Mirror and glass surfaces are essential components of our daily environment yet notoriously hard to scan. Starting from the simple idea of robustly detecting a reflected planar target, we demonstrate a complete system for robust and accurate reconstruction of scenes with mirrors and glass surfaces. Given the ease of capture, our system could also be used to collect training data for learning-based approaches to detect reflective surfaces. Besides our core application of scanning indoor scenes, we envision multiple extensions and applications.

First, our tag requires a relatively clear reflection in order to be detected by the AprilTag detector. Using different patterns and detectors, one could extend our method to glossy and specular surfaces. We also believe that our proposed technique could be extended to explicitly handle surfaces with curvature. Next, our tag provides a moving, active and patterned area light. We envision that this could be used to also infer reflectance information about other non-reflective surfaces in a scene. Finally, it would be interesting to evaluate whether and how our approach could be integrated into autonomous robots, allowing them to optically detect reflective surfaces, in particular when using only passive sensing techniques such as classical (multi-view) stereo.

## A DERIVATION OF DOUBLE IMAGES ON GLASS

Fig. 16 gives the geometry for a fronto-parallel scanning rig observing a glass pane with finite thickness. Given an incident ray hitting the glass surface at an angle  $\theta$ , the refracted ray inside the glass will travel under an angle  $\alpha$  as determined by Snell's law:

$$\frac{\sin \theta}{\sin \alpha} = \frac{n_{\text{glass}}}{n_{\text{air}}} \quad (11)$$

$n_{\text{glass}}$  and  $n_{\text{air}}$  are the indices of refraction of the materials. Given a baseline  $b$  between the camera and the tag and a distance  $d$  between the scanning rig and the glass, the offset  $D(d)$  can be computed as

$$D(d) = 2t \tan \alpha = 2t \tan \left( \sin^{-1} \left( \frac{n_{\text{air}}}{n_{\text{glass}}} \sin \theta \right) \right) \quad (12)$$

$$= 2t \tan \left( \sin^{-1} \left( \frac{n_{\text{air}}}{n_{\text{glass}}} \sin \left( \tan^{-1} \frac{b}{2d} \right) \right) \right) \quad (13)$$

Fig. 16 shows the offset  $D(d)$  for a glass thickness of 5 mm, a relative index of refraction  $\frac{n_{\text{air}}}{n_{\text{glass}}}$  of 0.66 and a baseline of 0.25 m, which corresponds approximately to our setup. A single pixel on our AprilTag is approximately 3.5 mm wide, which corresponds to the offset at roughly 0.2 m distance.

## REFERENCES

- Sameer Agarwal, Keir Mierle, and Others. 2018. Ceres Solver. <http://ceres-solver.org> (2018).
- N. Arvanitopoulos, R. Achanta, and S. Süstrunk. 2017. Single Image Reflection Suppression. In *CVPR 2017*. 1752–1760.
- J. Balzer, D. Acevedo-Feliz, S. Soatto, S. Höfer, M. Hadwiger, and J. Beyerer. 2014. Cavlectometry: Towards Holistic Reconstruction of Large Mirror Objects. In *2nd International Conference on 3D Vision (3DV)*. 448–455.
- J. Balzer, S. Höfer, and J. Beyerer. 2011. Multiview specular stereo reconstruction of large mirror surfaces. In *CVPR 2011*. 2537–2544.
- L. G. Brown. 1992. A Survey of Image Registration Techniques. *Computing Surveys* 24, 4 (December 1992), 325–376.
- Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. 2017. Matterport3D: Learning from RGB-D Data in Indoor Environments. In *5th International Conference on 3D Vision (3DV)*.
- Tongbo Chen, Michael Goesele, and Hans-Peter Seidel. 2006. Mesostructure from Specularity. In *CVPR 2006*, Vol. 2. 1825–1832.
- Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. 2017. ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE.
- A. DelPozo and S. Savarese. 2007. Detecting Specular Surfaces on Natural Images. In *CVPR 2007*.
- Yuanyuan Ding and Jingyi Yu. 2008. Recovering shape characteristics on near-flat specular surfaces. In *CVPR 2008*.
- J. Engel, V. Koltun, and D. Cremers. 2018. Direct Sparse Odometry. *PAMI* 40, 3 (2018), 611–625.
- A. Fasano, M. Callieri, P. Cignoni, and R. Scopigno. 2003. Exploiting mirrors for laser stripe 3D scanning. In *3DIM 2003*. 243–250.
- Paul Foster, Zhenghong Sun, Jong Jin Park, and Benjamin Kuipers. 2013. VisAGGE: Visible angle grid for glass environments. In *CVPR 2013*. 2213–2220.
- C. Godard, P. Hedman, W. Li, and G. J. Brostow. 2015. Multi-view Reconstruction of Highly Specular Surfaces in Uncontrolled Environments. In *3rd International Conference on 3D Vision (3DV)*. 19–27.
- Ivo Ihrke, Kiriaios N. Kutulakos, Hendrik P. A. Lensch, Marcus Magnor, and Wolfgang Heidrich. 2010. Transparent and Specular Object Reconstruction. *Computer Graphics Forum* 29, 8 (2010), 2400–2426.
- B. Jacquet, C. Häne, K. Köser, and M. Pollefeys. 2013. Real-World Normal Map Capture for Nearly Flat Reflective Surfaces. In *CVPR 2013*. 713–720.
- Jun Jiang, Renato Miyagisuku, Atsushi Yamashita, and Hajime Asama. 2017. Glass Confidence Maps Building Based on Neural Networks Using Laser Range-Finders for Mobile Robots. In *IEEE/SICE International Symposium on System Integration*.
- O. Kähler, V. Adrian Prisacariu, C. Yuheng Ren, X. Sun, P. Torr, and D. Murray. 2015. Very High Frame Rate Volumetric Integration of Depth Images on Mobile Devices. *TVCG* 21, 11 (Nov 2015), 1241–1250.
- J. Kannala and S. S. Brandt. 2006. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *PAMI* 28, 8 (Aug 2006), 1335–1340.
- P.-F. Kashammer and A. Nüchter. 2015. Mirror identification and correction of 3D point clouds. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40, 5 (2015), 109.
- U. Klank, D. Carton, and M. Beetz. 2011. Transparent object detection and reconstruction on a mobile platform. In *ICRA 2011*. 5971–5978.
- Rainer Koch, Stefan May, Patrick Murmann, and Andreas Nüchter. 2017b. Identification of Transparent and Specular Reflective Material in Laser Scans to Discriminate Affected Measurements for Faultless Robotic SLAM. *Journal of Robotics and Autonomous Systems (JRAS)* 87 (2017), 296–312.
- R. Koch, S. May, and A. Nüchter. 2017a. Effective distinction of transparent and specular reflective objects in point clouds of a multi-echo laser scanner. In *ICAR 2017*. 566–571.
- Brian Kulis and Michael I. Jordan. 2011. Revisiting k-means: New algorithms via Bayesian nonparametrics. *arXiv preprint arXiv:1111.0352* (2011).
- J. P. Lewis. 1995. Fast Normalized Cross-Correlation. In *Vision Interface '95*. Canadian Image Processing and Pattern Recognition Society.
- M. Liu, R. Hartley, and M. Salzmann. 2015. Mirror Surface Reconstruction from a Single Image. *PAMI* 37, 4 (April 2015), 760–773.
- William E. Lorensen and Harvey E. Cline. 1987. Marching cubes: A high resolution 3D surface construction algorithm. In *SIGGRAPH*. ACM, 163–169.
- D. Miyazaki, M. Kagesawa, and K. Ikeuchi. 2004. Transparent surface modeling from a pair of polarization images. *PAMI* 26, 1 (2004), 73–82.
- R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós. 2015. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics* 31, 5 (Oct 2015), 1147–1163.
- R. A. Newcombe, S. Izadi, O. Hilliges, D. Molnyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. 2011. KinectFusion: Real-time dense surface mapping and tracking. In *ISMAR 2011*. 127–136.
- Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Marc Stamminger. 2013. Real-time 3D Reconstruction at Scale Using Voxel Hashing. *ACM Trans. Graph.* 32, 6, Article 169 (Nov. 2013), 11 pages.
- Edwin Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *ICRA 2011*. 3400–3407.
- Rui Rodrigues, João P. Barreto, and Urbano Nunes. 2010. Camera Pose Estimation Using Images of Planar Mirror Reflections. In *ECCV 2010*. 382–395.
- YiChang Shih, D. Krishnan, F. Durand, and W. T. Freeman. 2015. Reflection removal using ghosting cues. In *CVPR 2015*. 3193–3201.
- Sudipta N. Sinha, Johannes Kopf, Michael Goesele, Daniel Scharstein, and Richard Szeliski. 2012. Image-based Rendering for Scenes with Reflections. *ACM Trans. Graph.* 31, 4, Article 100 (July 2012), 10 pages.
- Julian Straub, Trevor Campbell, Jonathan P. How, and John W. Fisher. 2015. Small-variance nonparametric clustering on the hypersphere. In *CVPR 2015*. 334–342.
- Marco Tarini, Hendrik P.A. Lensch, Michael Goesele, and Hans-Peter Seidel. 2005. 3D acquisition of mirroring objects using striped patterns. *Graphical Models* 67, 4 (2005), 233 – 259.
- Markus Unger, Thomas Pock, and Horst Bischof. 2008a. *Interactive globally optimal image segmentation*. Technical Report ICG-TR-08/02. Graz University of Technology.
- Markus Unger, Thomas Pock, Werner Trobin, Daniel Cremers, and Horst Bischof. 2008b. TVSeg - Interactive Total Variation Based Image Segmentation. In *BMVC 2008*.
- J. Wang and E. Olson. 2016. AprilTag 2: Efficient and robust fiducial detection. In *IROS 2016*. 4193–4198.
- Qiaosong Wang, Haiping Lin, Yi Ma, Sing Bing Kang, and Jingyi Yu. 2015. Automatic Layer Separation using Light Field Imaging. *CoRR* abs/1506.04721 (2015). arXiv:1506.04721 <http://arxiv.org/abs/1506.04721>
- Sven Wanner and Bastian Goldluecke. 2013. Reconstructing Reflective and Transparent Surfaces from Epipolar Plane Images. In *GCPR 2013*.
- Tianfan Xue, Michael Rubinstein, Ce Liu, and William T. Freeman. 2015. A computational approach for obstruction-free photography. *ACM Trans. Graph.* 34, 4 (2015), 79:1–79:11.
- Shao-Wen Yang and Chieh-Chih Wang. 2008. Dealing with laser scanner failure: Mirrors and windows. In *ICRA 2008*. 3009–3015.
- S. W. Yang and C. C. Wang. 2011. On Solving Mirror Reflection in LIDAR Sensing. *IEEE/ASME Transactions on Mechatronics* 16, 2 (April 2011), 255–265.
- Y. Zhang, M. Ye, D. Manocha, and R. Yang. 2017. 3D Reconstruction in the Presence of Glass and Mirrors by Acoustic and Visual Fusion. *PAMI* (2017).
- Zhengyou Zhang. 2000. A Flexible New Technique for Camera Calibration. *PAMI* 22, 11 (Nov. 2000), 1330–1334.