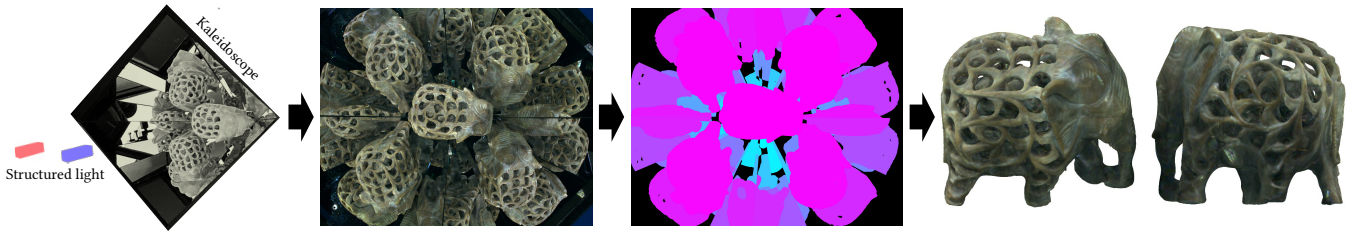
Kaleidoscopic Structured Light

BYEONGJOO AHN, IOANNIS GKIOULEKAS, and ASWIN C. SANKARANARAYANAN,
Carnegie Mellon University, USA



(a) Imaging setup (b) Scanning (c) Labeling (d) Full surround 3D reconstruction

Fig. 1. **Kaleidoscopic structured light.** (a) We propose a system for full surround 3D imaging using an imaging setup that consists of a projector, a camera, and a kaleidoscope. (b) The camera and projector observe the object from a large number of virtual viewpoints, which unravel the complex geometry of the object. (c) To use the kaleidoscopic image for multi-view stereo, we label each pixel, i.e., identify the sequence of mirrors that the ray backprojected from the pixel encounters before intersecting the object. (d) This labeling allows us to reconstruct the shape of the object using multi-view triangulation. Our system enables us to reconstruct, with high accuracy and full coverage, highly complex objects that have intricate geometric features, including concavities and self-occlusions. Reconstructed point clouds and full surround videos are available on the project webpage [Ahn et al. 2021].

Full surround 3D imaging for shape acquisition is essential for generating digital replicas of real-world objects. Surrounding an object we seek to scan with a kaleidoscope, that is, a configuration of multiple planar mirrors, produces an image of the object that encodes information from a combinatorially large number of virtual viewpoints. This information is practically useful for the full surround 3D reconstruction of the object, but cannot be used directly, as we do not know what virtual viewpoint each image pixel corresponds—the pixel label. We introduce a structured light system that combines a projector and a camera with a kaleidoscope. We then prove that we can accurately determine the labels of projector and camera pixels, for arbitrary kaleidoscope configurations, using the projector-camera epipolar geometry. We use this result to show that our system can serve as a multi-view structured light system with hundreds of virtual projectors and cameras. This makes our system capable of scanning complex shapes precisely and with full coverage. We demonstrate the advantages of the kaleidoscopic structured light system by scanning objects that exhibit a large range of shapes and reflectances.

CCS Concepts: • **Computing methodologies** → **3D imaging**; **Computational photography**; **Epipolar geometry**.

Additional Key Words and Phrases: structured light, kaleidoscope, epipolar geometry, multi-view stereo

ACM Reference Format:
Byeongjoo Ahn, Ioannis Gkioulekas, and Aswin C. Sankaranarayanan. 2021. Kaleidoscopic Structured Light. *ACM Trans. Graph.* 0, 0, Article 0 (2021), 15 pages. https://doi.org/10.1145/1122445.1122456

Authors' address: Byeongjoo Ahn, bahn@cmu.edu; Ioannis Gkioulekas, igkioule@andrew.cmu.edu; Aswin C. Sankaranarayanan, saswin@andrew.cmu.edu
, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA.

1 INTRODUCTION

3D scanning of a single view of an object seldom suffices. Be it for 3D printing, augmented reality, or virtual reality, scanning of the shape of the entire object in all its complexity—what we refer to as *full surround 3D*—is critical to have a faithful digital twin.

A key factor in achieving a full surround 3D scan is the number of viewpoints from which an object is imaged. Covering the entire object that we seek to scan typically requires a large number of diverse viewpoints. Furthermore, this number increases with the complexity of the object. This requirement has led to the construction of light stages with multiple cameras, and potentially projectors, that capture digital content at high fidelity. Unfortunately, the cost and complexity of these systems place them beyond the reach of the average consumer of 3D technology.

A simple way to achieve a very large number of viewpoints is to surround the object we want to scan with mirrors, which conveniently provide additional viewpoints without the need to move a camera or construct a multiple-camera system. In particular, a kaleidoscope [Brewster 1858], which consists of multiple planar mirrors, allows light to bounce around repeatedly until it hits the camera, thereby providing a combinatorial increase in the number of viewpoints. Thus, with a kaleidoscope and a single camera, we can construct a *virtual* multi-view imaging system that is easy to build, calibrate and deploy, with components that are easily available.

The key challenge when using a kaleidoscope, however, lies in interpreting the captured image and decoding the numerous views of the object that it provides. Specifically, we need to identify the virtual viewpoint corresponding to each pixel on the captured image. We call this the *labeling* problem. The labeling information allows us to decompose the single captured image into multiple segments, one for each virtual viewpoint. In the absence of this information, we cannot estimate the 3D shape by triangulating from correspondences across different views. The fact that in a kaleidoscope it is common

Table 1. **Comparison of surround reconstruction methods.**

Method	Registration	Correspondence	Labeling	Cost
Single camera	x	x	-	low
Multi-camera system	o	x	-	high
Multi-projector camera system	o	o	-	very high
Camera + mirrors [Reshetouski et al. 2011]	o	x	visual hull	low
ToF + mirrors [Xu et al. 2018]	o	o	path length	high
Structured light + mirrors [Lanman et al. 2009]	o	o	manual	medium
Structured light + mirrors (ours)	o	o	epipolar constraint	medium

to observe hundreds of virtual views that are interwoven with each other makes the labeling problem particularly challenging.

We propose a full surround 3D imaging system that we call kaleidoscopic structured light, comprising a projector, a camera, and a kaleidoscope. Our main technical result is to show that we can correctly label the virtual projectors and virtual cameras for arbitrary kaleidoscope configurations, by using their *epipolar geometry* and other physical constraints arising from image formation for this setup. With this result, our kaleidoscopic structured light system can serve as a multi-view structured light system with tens, if not hundreds, of virtual projectors and cameras, which are hard to construct with real devices. Our system allows us to reconstruct shape (in the form of triangular meshes) of highly complex objects with intricate features, by providing correspondences from multiple viewpoints with full surround coverage.

Contributions. Our work advances the state of the art of 3D scanning, by facilitating the reconstruction of challenging objects, such as textureless or glossy objects, with complex 3D geometry. This is made possible through the following contributions.

- *Labeling using epipolar geometry.* We develop a labeling technique that takes as input correspondences between a projector pixel illuminating a point on the object, and multiple camera pixels observing that point. Our labeling technique uses constraints specific to the kaleidoscopic setup, and in particular the epipolar geometry across these multiple correspondences, to jointly decode the mirror sequences at the projector and camera pixels.

- *Theoretical guarantees on the recoverability of labels using epipolar geometry.* We prove a theoretical result that establishes the uniqueness and correctness of the labels we decode with our labeling technique. Specifically, we show that the labels we recover from the projector-camera pixel correspondences are accurate, provided the epipolar planes corresponding to the projector-camera geometry are not parallel to the mirror normals. This condition is easy to achieve by placing the camera and projector in asymmetric poses relative to the kaleidoscope.

- *Fast scanning techniques.* We develop heuristics for speeding up our scan, exploiting the fact that the labeling problem is highly over-constrained due to multiple camera pixels mapping to the same projector pixel.

These advances in kaleidoscopic imaging allow us to push the boundaries of 3D scanning in terms of the complexity of the objects that can be reconstructed. Fig. 1 shows an example 3D scan of an object that fully encompasses another smaller object, and is characterized by concavities, a large genus number, and complex self-occlusion. We have released our code and data on the project website [Ahn et al. 2021], to facilitate reproducibility and follow-up research.

Limitations. The main limitation of the proposed kaleidoscopic structured light technique is the time it requires for scanning an object. The baseline version of our technique relies on point scanning, which can take a prohibitive amount of time. We provide parallelized scanning techniques to speed up acquisition, but our scan times remain considerably longer than those achieved by traditional structured light acceleration techniques. Unfortunately, the complex nature of kaleidoscopic light transport makes these traditional techniques inapplicable to our setting.

2 RELATED WORK

We review prior work on full surround 3D reconstruction. We summarize key features of our and prior techniques in Table 1.

2.1 Full Surround 3D Reconstruction

Common approaches for reconstructing a complete scan of an object include rotating the object [Kang et al. 2019; NextEngine 2000; Park et al. 2016; Xia et al. 2016; Zhou et al. 2013], moving a camera around the object [Cui et al. 2010; Holroyd et al. 2010; Kolev et al. 2014; Lichy et al. 2021; Nam et al. 2018; Newcombe et al. 2011; Ondrúška et al. 2015; Wu et al. 2015; Wu and Zhou 2015], or constructing a multi-camera system [Ghosh et al. 2011; Joo et al. 2017; Schwartz et al. 2013]. However, rotating the object around a fixed axis (e.g., using a turntable) constrains viewpoints to lie on a plane perpendicular to the rotation axis; such viewpoints may be insufficient for the full surround reconstruction of intricate objects. Moving a single camera requires estimating the camera pose, whereas multi-camera systems are generally costly and difficult to build. These approaches are even more challenging when we try to scan textureless objects, for which correspondences are hard to establish using only passive illumination. Even though projector-camera systems can help solve the correspondence problem, such systems are even more difficult to scale to the full surround case, in part due to the complexity and cost, and in part due to interference between multiple projectors.

A fascinating approach for full surround 3D reconstruction is the so-called *dip transform*, where the object is dipped into a fluid at

multiple orientations [Aberman et al. 2017]. Measuring the liquid displacement, which encodes the submerged volume, provides sufficient information to recover the shape of the object. Although this method provides superior results for intricate objects, it requires immersing the object in the fluid, which is not always feasible.

2.2 3D Reconstruction with Mirrors

Label-free approaches. There is extensive literature on the use of mirrors for 3D reconstruction, in combination with passive-illumination camera systems [Forbes et al. 2006; Fuchs et al. 2013; Gluckman and Nayar 2001, 2002; Goshtasby and Gruver 1993; Hu et al. 2005; Huang and Lai 2006; Mitsumoto et al. 1992; Murray 1995; Nene and Nayar 1998; Taguchi et al. 2010a,b; Ying et al. 2010], time-of-flight (ToF) cameras [Nobuhara et al. 2016], and projector-camera systems [Bangay and Radloff 2004; Garg et al. 2006; Han and Perlin 2003; Lanman et al. 2009; Tahara et al. 2015]. The use of mirrors is in large part due to the increase in viewpoints they provide. In most systems using mirrors this way, the number of reflections and virtual viewpoints are carefully controlled to be few in number, which makes manual labeling practical.

Lanman et al. [2009] propose a structured light system that combines an orthographic projector, a camera, and mirrors, and is designed to remove the interference between reflected projector patterns. They solve the interference problem by illuminating with patterns that are perfectly aligned after one or multiple mirror reflections. However, achieving this requires using a special configuration of mirrors, which in turn constrains the locations of virtual cameras to lie on a plane. Such a viewpoint set can be insufficient for full surround 3D coverage when scanning complex objects. Additionally, their configuration has only four virtual cameras, to make manual labeling of the virtual images observed in the camera practical. Tahara et al. [2015] extend this approach to perspective projectors, again with manual labeling. Kaleidoscopes have also been used for measuring bidirectional texture functions [Bangay and Radloff 2004; Han and Perlin 2003]; in these works, the underlying shapes are nearly planar, which simplifies solving the labeling problem.

Approaches that estimate labels. Using numerous virtual viewpoints for better full surround coverage requires being able to estimate labels automatically. Reshetouski et al. [2011] solve the labeling problem in a passive-illumination kaleidoscopic system by using space carving [Kutulakos and Seitz 2000]. As the background pixels on the captured image do not intersect with the object, even after repeated mirror reflections, the rays corresponding to those pixels can be backprojected and "carved". This provides a visual hull of the object inside the kaleidoscope, which can be combined with ray tracing to obtain the label map. Ihrke et al. [2012] combine this labeling method with a structured light system that illuminates only one label at a time, to obtain correspondences while avoiding interference. As these methods rely on space carving for labeling, they are inaccurate when the visual hull is not a good approximation to the object; typically, this is the case for concave objects. Xu et al. [2018] combine a kaleidoscope with a ToF camera. The ToF information provides the total path length from the camera to the object. This allows estimating the label and 3D point, by folding the ray using the known mirror configuration. Unfortunately, this approach requires using a pulsed ToF system, which can be costly.

3 OVERVIEW

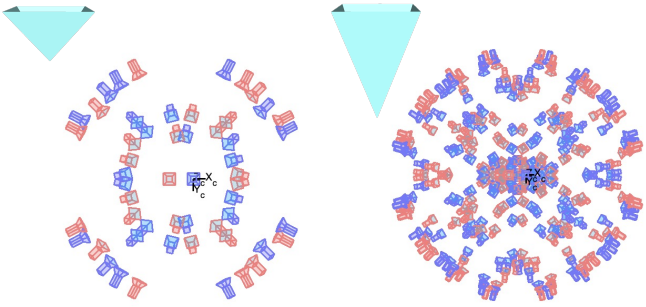
We provide an overview of the proposed kaleidoscopic structured light system and its associated reconstruction pipeline.

3.1 Imaging Setup

Hardware. We construct the kaleidoscopic structured light system by combining a projector, a camera, and a kaleidoscope. We create the kaleidoscope by placing four planar mirrors in a pyramidal shape, which Xu et al. [2018] report empirically to provide good scanning coverage. We place the projector and camera at the bottom of the pyramid, oriented to look at its tip. We calibrate the intrinsic parameters of the projector and camera, and the extrinsic pose of the projector and mirrors relative to the camera. Then, we place the object inside the mirror system, either by hanging it with strings, or by placing it on the mirrors directly. The kaleidoscopic arrangement provides hundreds of virtual projectors and cameras. We show a schematic of our setup in Fig. 1, and provide details about our hardware prototype and calibration procedure in Section 6. Fig. 1 also outlines the steps of the imaging pipeline—scanning, labeling, and shape reconstruction—which we review next.

Scanning. We first scan the object by turning "on" a number of projector pixels, which illuminate a set of object locations, and thus camera pixels. To simplify exposition, we describe our approach assuming that we activate only a single projector pixel at a time. In Section 6, we discuss techniques for speeding up acquisition by simultaneously activating multiple projector pixels. The projector pixel that we activate illuminates a single point on the object surface, either directly or after one or multiple reflections on the kaleidoscope. The illuminated point is observed at multiple camera pixels, each via a different sequence of mirror reflections, and we save the locations of these pixels. This provides us with a correspondence between a single projector pixel and multiple camera pixels, all of which also correspond to a single object point. This interpretation assumes that interreflections on the object are weak enough to not overwhelm the direct observation of the illuminated point; in our experience, this is generally true except for when the scanned object is a mirror or highly-specular.

Labeling. We now have a correspondence between a single projector pixel and multiple camera pixels. However, we cannot directly use this correspondence for triangulation, because we do not know the mirror sequence encountered by the rays corresponding to the projector and camera pixels. For each projector or camera pixel, we refer to the sequence of mirrors that the ray from the pixel encounters before intersecting the object as the pixel's *label* [Reshetouski et al. 2011]. We refer to the task of determining the labels for all projector and camera pixels as the *labeling problem*. Solving the labeling problem is the core challenge of kaleidoscopic imaging. Labeling and shape reconstruction are interwoven, as labeling requires reasoning about the object shape, and shape reconstruction requires using the labels to perform triangulation. We will define



(a) Wide pyramid (b) Sharp pyramid

Fig. 2. **Mirror design.** Virtual projectors and cameras produced by pyramidal kaleidoscopes with different tip angles. A sharp pyramid produces more virtual projectors and cameras, which can provide better coverage. The angles of pyramid tips are (a) 72° and (b) 36°, and the number of the virtual cameras (or projectors) are 27 and 145, respectively.

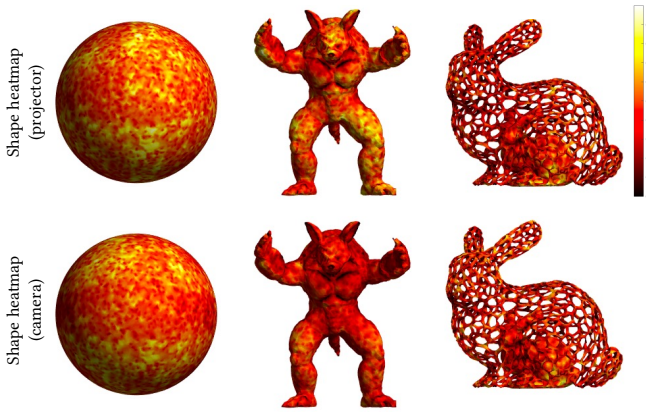


Fig. 3. **Surface coverage.** Our kaleidoscopic structured light system generates hundreds of virtual projectors and cameras, which densely sample the light-view space. The heatmaps visualize the number of camera and projector pixels that intersect each surface facet (normalized in each object).

the labeling problem more formally in Section 3.3, and show how to solve it using epipolar geometry constraints in Section 4.

Shape reconstruction. The projector-camera correspondences and labeling information allow us to reconstruct 3D geometry using multi-view triangulation: For each projector pixel, we use its correspondence with multiple camera pixels to estimate a 3D point that comes closest to intersecting all pixel rays (Section 5).

3.2 Kaleidoscope Design and Coverage

We use the pyramid mirror configuration from Xu et al. [2018], with one modification: We adjust the angle of the pyramid to increase the number of virtual viewpoints. Having more viewpoints is important for improving the accuracy of the multi-view triangulation procedure our technique uses. This is different from the methodology of Xu et al. [2018], where the ToF depth-sensing mechanism does not require having multiple views of the same point. Having more views also helps improve coverage, and thus can enable reconstructing occluded parts of objects with highly complex visibility. Fig. 2 shows that the sharp pyramid provides significantly more virtual projectors and cameras than one with a larger angle. Our new configuration also provides good scanning coverage, as we show empirically in Fig. 3, where we visualize the number of projector and camera pixels that observe each vertex of an object mesh. We leave the systematic optimization of the kaleidoscope configuration as an important future research direction.

3.3 Basics of Kaleidoscopic Imaging

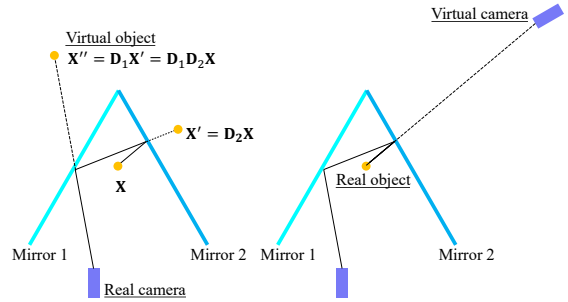
Before we introduce how to solve the labeling problem, we review the transformation of rays and points by planar mirrors, and the epipolar geometry between the virtual projectors and cameras.

Transformation by planar mirrors. To represent the transformation by a single planar mirror m, we define the mirror as a plane with normal $\mathbf{n}$ and distance from origin d (represented in world coordinates). Then, the 4×4 reflection matrix in homogeneous coordinates can be written as

$$\mathbf{D}_m = \begin{bmatrix} \mathbf{I} - 2\mathbf{n}\mathbf{n}^\top & 2d\mathbf{n} \\ \mathbf{0} & 1 \end{bmatrix}. \tag{1}$$

Note that $\mathbf{D}_m$ is an involutory matrix (i.e., $\mathbf{D}_m^2 = \mathbf{I}$), as the reflection of a reflected point is the same as the original point.

The transformation by multiple planar mirrors can be represented by multiplying the reflection matrices $\{\mathbf{D}_m\}$ corresponding to each mirror. Consider the example of Fig. 4: a ray from a 3D point $\mathbf{X}$ bounces off mirrors 2 and 1 before being observed at a camera pixel, as shown in Fig. 4. Then, the intermediate virtual point after the first reflection at mirror 2 equals $\mathbf{X}' = \mathbf{D}_2\mathbf{X}$, and the final virtual point seeing $\mathbf{X}'$ through mirror 1 equals $\mathbf{X}'' = \mathbf{D}_1\mathbf{X}' = \mathbf{D}_1\mathbf{D}_2\mathbf{X}$.



(a) Virtual-object interpretation (b) Virtual-camera interpretation

Fig. 4. **Mirror transformation interpretations.** We can interpret the mirror transformation in two mathematically equivalent ways: (a) *Virtual object*—the location of the *virtual* point in the *real* camera coordinate system, corresponding to unfolding the ray from the object side. (b) *Virtual camera*—the location of the *real* point in the *virtual* camera coordinate system, corresponding to unfolding the ray from the camera side.

Labels. As mentioned earlier, successful shape recovery requires us to decode the light path between a scene point and the projector or camera pixel observing it. Following Reshetouski et al. [2011],

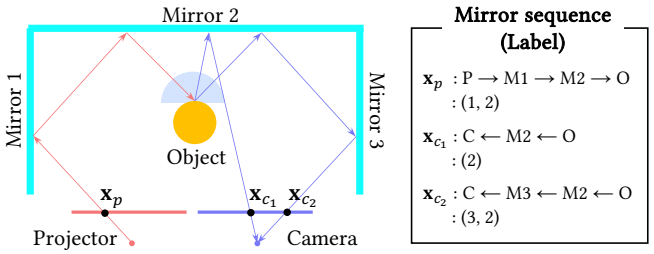


Fig. 5. **Label definition.** The label of a projector or camera pixel is the sequence of mirrors that the ray backprojected from the pixel reflects off of before reaching the scanned object.

we represent this path using a sequence of mirror labels; for a kaleidoscope with M mirrors, and a light path with K reflections, the label sequence is:

$$\ell \equiv (l_k)_{k=1}^{K} = (l_1, l_2, l_3, \ldots, l_K), \tag{2}$$

where $l_k \in \{1, 2, ..., M\}$ denotes a mirror label. We use the convention that the labeling starts at the pixel and ends at the object. For instance, in Fig. 5, the number of mirrors are $M = 3$, the labels for the light path shown are $\ell_p = (1, 2)$, $\ell_{c_1} = (2)$, and $\ell_{c_2} = (3, 2)$. With this label definition, the mirror transformation matrix $\mathbf{D}(\ell)$ can be written as a function of the label ℓ as

$$\mathbf{D}(\ell) = \prod_{k=1}^{K} \mathbf{D}_{l_k} = \mathbf{D}_{l_1}\mathbf{D}_{l_2}\mathbf{D}_{l_3}\cdots\mathbf{D}_{l_K}. \tag{3}$$

Virtual cameras and projectors. As Fig. 4 shows, the transformed 3D point $\mathbf{X}'' = \mathbf{D}\mathbf{X}$ can be interpreted in two ways: (i) *virtual object*—the location of the *virtual* point in the *real* camera coordinate system; or (ii) *virtual camera*—the location of the *real* point in the *virtual* camera coordinate system. The first interpretation corresponds to unfolding the ray from the object side, and the second to unfolding the ray from the pixel side. Each of these equivalent interpretations has its own advantages. For example, the virtual-object interpretation is useful when we derive the transformation by multiple mirrors, as it lets us avoid the change of camera coordinates. By contrast, the virtual-camera interpretation is useful when we derive expressions for triangulated 3D points.

We can use the mathematical equivalence of these two interpretations to represent the extrinsic matrix of the virtual camera $\mathbf{T}_{\text{virtual}}$: Let $\mathbf{T}_{\text{real}}$ be the 4×4 extrinsic matrix of the real camera (i.e., camera pose in the world coordinate system). Then, the virtual 3D point $\mathbf{X}_{\text{virtual}} = \mathbf{D}\mathbf{X}_{\text{real}}$ in the real camera coordinate system equals

$$\mathbf{T}_{\text{real}}\mathbf{X}_{\text{virtual}} = \mathbf{T}_{\text{real}}\mathbf{D}\mathbf{X}_{\text{real}} = \mathbf{T}_{\text{virtual}}\mathbf{X}_{\text{real}}. \tag{4}$$

This can be interpreted as a real 3D point in the local coordinate system of a virtual camera whose extrinsic matrix equals

$$\mathbf{T}_{\text{virtual}} = \mathbf{T}_{\text{real}}\mathbf{D}. \tag{5}$$

The above derivation applies, mutatis mutandis, to projectors, and their virtual counterparts. We skip the derivation as, for the most part, projectors can be treated as cameras.

Epipolar geometry of virtual projectors and cameras. Now we can represent the virtual projectors and cameras using the parameters that define the real projector, camera, and mirrors. Let the extrinsic matrix of the real projector and camera be $\mathbf{T}_p$ and $\mathbf{T}_c$, respectively. Then, the extrinsic matrix of the virtual projector with mirror transformation $\mathbf{D}_p$ is $\mathbf{T}'_p = \mathbf{T}_p\mathbf{D}_p$, and that of the virtual camera with mirror transformation $\mathbf{D}_c$ is $\mathbf{T}'_c = \mathbf{T}_c\mathbf{D}_c$. Thus, a 3D scene point $\mathbf{X}$ observed from the virtual projector and the virtual camera equals

$$\begin{cases} \mathbf{X}_p = \mathbf{T}'_p\mathbf{X} = \mathbf{T}_p\mathbf{D}_p\mathbf{X}, \\ \mathbf{X}_c = \mathbf{T}'_c\mathbf{X} = \mathbf{T}_c\mathbf{D}_c\mathbf{X}, \end{cases} \tag{6}$$

where $\mathbf{X}_p$ and $\mathbf{X}_c$ are the representations of the point in the local coordinates of the virtual projector and virtual camera, respectively.

To mathematically describe the epipolar geometry between the virtual projector and virtual camera, we can express the relative transformation between them as

$$\mathbf{X}_p = \mathbf{T}_p\mathbf{D}_p\mathbf{X} = \mathbf{T}_p\mathbf{D}_p\mathbf{D}_c^{-1}\mathbf{T}_c^{-1}\mathbf{X}_c = \begin{bmatrix}\mathbf{R} & \mathbf{t}\\ \mathbf{0} & 1\end{bmatrix}\mathbf{X}_c, \tag{7}$$

where $\mathbf{R}$ and $\mathbf{t}$ are the relative rotation and translation between the virtual projector and virtual camera, respectively. Finally, to express the fundamental matrix, we denote by $\mathbf{K}_p$ and $\mathbf{K}_c$ the intrinsic matrices of the virtual projector and virtual camera, respectively (i.e., $\mathbf{x}_p \sim \mathbf{K}_p\mathbf{X}_p$, and $\mathbf{x}_c \sim \mathbf{K}_c\mathbf{X}_c$). Then, we can express epipolar constraints between the virtual projector and virtual camera using the fundamental matrix $\mathbf{F} = \mathbf{K}_p^{-\top}[\mathbf{t}]_\times\mathbf{R}\mathbf{K}_c^{-1}$, which satisfies $\mathbf{x}_p^\top\mathbf{F}\mathbf{x}_c = 0$.

4 LABELING

We now present our main technical results, detailing the conditions under which the label sequence can be uniquely determined using the epipolar constraints between virtual projectors and cameras.

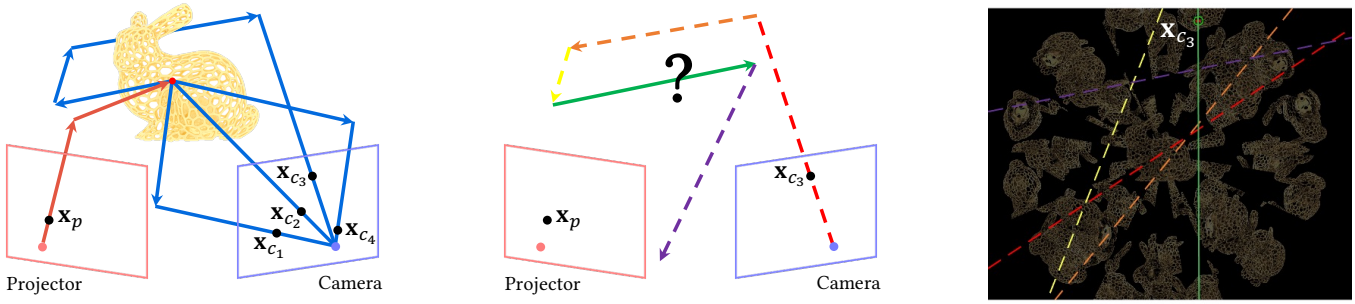
4.1 Problem Setup

Suppose that we illuminate a pixel $\mathbf{x}_p$ on the projector and observe a set of pixels $\{\mathbf{x}_{c_i}\}$ on the camera. Then, we formulate the labeling problem as follows: given the correspondence of $\mathbf{x}_p \leftrightarrow \{\mathbf{x}_{c_i}\}$, find the label, or mirror sequence, ℓ_p associated with the projector pixel $\mathbf{x}_p$, and the label set $\{\ell_{c_i}\}$ for each of the camera pixels in $\{\mathbf{x}_{c_i}\}$.

We approach this problem as one of finding label for the projector pixel and each of the camera pixels, such that the projector-camera pixel correspondences satisfy the epipolar constraints implied by the labels. Let the mirror transformation of the projector pixel $\mathbf{x}_p$ be $\mathbf{D}_p = \mathbf{D}(\ell_p)$, and that of the camera pixel $\mathbf{x}_{c_i}$ be $\mathbf{D}_{c_i} = \mathbf{D}(\ell_{c_i})$, as defined in (3). Then, the relative transformation between the virtual projector and camera is $\mathbf{T}_p\mathbf{D}_p\mathbf{D}_{c_i}^{-1}\mathbf{T}_c^{-1} = \begin{bmatrix}\mathbf{R}_i & \mathbf{t}_i\\ \mathbf{0} & 1\end{bmatrix}$. The fundamental matrix can now be written as a function of the labels ℓ_p and ℓ_{c_i} as $\mathbf{F}(\ell_p, \ell_{c_i}) = \mathbf{K}_p^{-\top}[\mathbf{t}_i]_\times\mathbf{R}_i\mathbf{K}_c^{-1}$. The epipolar distance due to the labels ℓ_p, ℓ_{c_i} and the correspondence $\mathbf{x}_p \leftrightarrow \mathbf{x}_{c_i}$ can be defined as

$$d(\ell_p, \ell_{c_i}; \mathbf{x}_p, \mathbf{x}_{c_i}) \equiv |\mathbf{x}_p^\top\mathbf{F}(\ell_p, \ell_{c_i})\mathbf{x}_{c_i}|. \tag{8}$$

We refer to (8) as the *virtual epipolar distance*, and our goal is to find projector and camera labels such that the virtual epipolar distance is zero for each of the projector-camera correspondences. Note that all the correspondences share the same projector label. In the presence



(a) Correspondence from scanning (b) Labeling problem (c) Epipolar labeling

Fig. 6. **Epipolar labeling.** We label how each pixel is reflected using epipolar labeling. (a) We obtain correspondences between a projector pixel and multiple camera pixels during scanning. (b) Labeling such a correspondence is equivalent to determining the number of reflections for each projector and camera pixel given the mirror geometry. We visualize the possible labels for the pixel $\mathbf{x}_{c_3}$. (c) We show on the camera image the epipolar lines corresponding to the possible labels, using matching colors. The green line passes through the pixel $\mathbf{x}_{c_3}$, which satisfies the epipolar constraint, whereas other lines do not. We prove that our epipolar labeling method can correctly determine the label for generic mirror configurations.

of noisy correspondences, we seek to minimize the total virtual epipolar distance given as

$$\min_{\ell_p, \{\ell_{c_i}\}} \sum_i d(\ell_p, \ell_{c_i}; \mathbf{x}_p, \mathbf{x}_{c_i}). \tag{9}$$

4.2 Epipolar Labeling

Optimizing (9) is not trivial, because there are exponentially many possible labels for both projector and camera; hence, exhaustive search is computationally intractable. The size of the search space for each projector or camera pixel is $O(M^{K_{\max}})$, where M is the number of mirrors and $K_{\max}$ is the maximum number of possible reflections in the kaleidoscope. In our prototype, $K_{\max} = 10$ (calculated by ray tracing using the calibration results). However, we can reduce the search space using the fact that the true label at a projector or camera pixel is a pre-subsequence of the label when there is no object in the kaleidoscope. In particular, we can precompute a pixel's *empty label* for an object-free kaleidoscope simply by ray tracing using the calibration results. The ray will be repeatedly reflected by the mirrors until it escapes the kaleidoscope. When we put an object in the kaleidoscope, the ray is truncated when it intersects the object. Thus, if we denote the empty label of a pixel $\mathbf{x}$ as

$$\ell^{\text{empty}}(\mathbf{x}) = (l_k)_{k=1}^{K_{\max}} = (l_1, l_2, l_3, \ldots, l_{K_{\max}}), \tag{10}$$

the label with an object will be a pre-subsequence of $\ell^{\text{empty}}(\mathbf{x})$,

$$\ell = (l_k)_{k=1}^{K} = (l_1, l_2, l_3, \ldots, l_K) \subset \ell^{\text{empty}}(\mathbf{x}). \tag{11}$$

Then, given the empty label, the labeling problem reduces to finding for each pixel the number of mirror reflections till intersecting the object. Let the number of reflections for $\mathbf{x}_p$ and $\mathbf{x}_{c_i}$ be K_p and K_{c_i}, respectively. We can rewrite (9) as

$$\min_{K_p, \{K_{c_i}\}} \sum_i d(\ell_p, \ell_{c_i}; \mathbf{x}_p, \mathbf{x}_{c_i}). \tag{12}$$

Now the search space size for each pixel reduces from $O(M^{K_{\max}})$ to $O(K_{\max})$, and we can solve the optimization problem (12) by linearly searching the number of reflections for $\mathbf{x}_p$ and $\{\mathbf{x}_{c_i}\}$. Importantly, the fact that this search couples a single projector pixel label with multiple camera pixel labels greatly enhances the robustness of the labeling procedure. Fig. 6 shows an example of epipolar labeling for one camera pixel, where our procedure finds the label that minimizes the virtual epipolar distance from the empty label.

4.3 Correctness of Epipolar Labeling

In Section 4.2, we described an efficient procedure for estimating labels for projector-camera pixel correspondences. We now analyze the correctness of this labeling procedure. Our analysis establishes two key facts: First, minimizing (12) determines labels up to a certain ambiguity. Second, resolving this ambiguity is possible by simply using the fact that the scanned object must lie in the physical space enclosed by the mirrors in front of the projector-camera system. Along the way, we derive conditions on the projector-camera-mirror geometry necessary for the correctness of our labeling procedure.

We define a *mirrored label* $M(\ell, m)$ of a label $\ell = (l_1, l_2, \ldots)$ as

$$M(\ell, m) \equiv (l_1, l_2, \ldots, m), \tag{13}$$

which is the label followed by an additional reflection by mirror m. We can generalize this to a multiply mirrored label, that is

$$M(\ell, \ell') \equiv (l_1, l_2, \ldots, l'_1, l'_2, \ldots), \tag{14}$$

This mirrored label is useful because all possible labels can be represented as mirrored labels of the true label. In particular, let the true label for $\mathbf{x}_p$ and $\mathbf{x}_{c_i}$ be ℓ_p^* and $\ell_{c_i}^*$, respectively, that is, $d(\ell_p^*, \ell_{c_i}^*; \mathbf{x}_p, \mathbf{x}_{c_i}) = 0$. Given that all possible labels are also pre-subsequence of the empty label, we can represent them as

$$\ell_p = M(\ell_p^*, \ell'_p), \quad \ell_c = M(\ell_{c_i}^*, \ell'_c), \tag{15}$$

where ℓ'_p and ℓ'_c are arbitrary labels contained in the empty label. We can now prove the following proposition.

Proposition 1 (Virtual epipolar distance of identically mirrored label.). *The virtual epipolar distance between ℓ_p and ℓ_c is the same as that between $M(\ell_p, \ell')$ and $M(\ell_c, \ell')$ for any label ℓ'.*

Proof. From (3), we can write the mirror transformation matrix of the mirrored label $\mathbf{D}(M(\ell, \ell'))$ as

$$\mathbf{D}(M(\ell, \ell')) = \mathbf{D}(\ell)\mathbf{D}(\ell'). \tag{16}$$

Then, the relative transformation between identically mirrored virtual projector and virtual camera becomes

$$\mathbf{T}_p\mathbf{D}(M(\ell_p, \ell'))\mathbf{D}(M(\ell_{c_i}, \ell'))^{-1}\mathbf{T}_c^{-1} \tag{17}$$

$$=\mathbf{T}_p\mathbf{D}(\ell_p)\mathbf{D}(\ell')\mathbf{D}(\ell')^{-1}\mathbf{D}(\ell_{c_i})^{-1}\mathbf{T}_c^{-1} \tag{18}$$

$$=\mathbf{T}_p\mathbf{D}(\ell_p)\mathbf{D}(\ell_{c_i})^{-1}\mathbf{T}_c^{-1}. \tag{19}$$

Therefore, the effect of the additional reflection cancels out, and the relative transformation does not change by the identically mirrored label. Thus, the epipolar distance does not change either. □

Remark. Proposition 1 implies that, given the true projector-camera label pair ℓ_p^* and $\ell_{c_i}^*$, any mirrored label pair of the form $M(\ell_p^*, \ell')$ and $M(\ell_{c_i}^*, \ell')$ satisfies

$$d(M(\ell_p^*, \ell'), M(\ell_{c_i}^*, \ell'); \mathbf{x}_p, \mathbf{x}_{c_i}) = d(\ell_p^*, \ell_{c_i}^*; \mathbf{x}_p, \mathbf{x}_{c_i}) = 0. \tag{20}$$

Therefore, given the true label, mirroring both the projector and the camera pixel with the same sequence produces a valid solution. This raises the question: can there be other ambiguous solutions to the labeling problem? That is, could a differently mirrored label be used for the projector and camera and still satisfy the epipolar constraints? We eliminate this possibility next.

Proposition 2 (Virtual epipolar distance of differently mirrored label.). *Given the true labels, ℓ_p^* and $\ell_{c_i}^*$, for the projector and camera pixels, the rays corresponding to the mirrored labels $M(\ell_p^*, \ell_p')$ and $M(\ell_{c_i}^*, \ell_c')$ never meet for $\ell_p' \neq \ell_c'$, i.e,*

$$d(M(\ell_p^*, \ell_p'), M(\ell_{c_i}^*, \ell_c'); \mathbf{x}_p, \mathbf{x}_{c_i}) > 0, \tag{21}$$

provided that the kaleidoscope and the projector-camera pair are in a generic configuration where the epipolar planes, both real and virtual, and mirror normals are not co-planar.

Proof. Our proof relies on the intuition that the probability of two arbitrary lines in 3D being co-planar is zero. We explain the proof when the mirrored label introduces a single additional bounce. Without loss of generality, we assume that the true 3D point is at the origin, and that the last two mirror bounces before hitting the object are at points $\mathbf{p}_1$ and $\mathbf{p}_2$, on mirrors 1 and 2, respectively. If we have the true labels, the rays for triangulation are $\mathbf{p}_1 + t_1(-\mathbf{p}_1)$ and $\mathbf{p}_2 + t_2(-\mathbf{p}_2)$ and intersect at the origin. If we have the wrong labels, which include additional bounces on mirrors 1 and 2, the rays are $\mathbf{p}_1 + t_1(\mathbf{I} - 2\mathbf{n}_1\mathbf{n}_1{}^\top)\mathbf{p}_1$ and $\mathbf{p}_2 + t_2(\mathbf{I} - 2\mathbf{n}_2\mathbf{n}_2{}^\top)\mathbf{p}_2$. The two rays are co-planar if and only if

$$\det\left(\begin{bmatrix}\mathbf{p}_1 - \mathbf{p}_2 & (\mathbf{I} - 2\mathbf{n}_1\mathbf{n}_1{}^\top)\mathbf{p}_1 & (\mathbf{I} - 2\mathbf{n}_2\mathbf{n}_2{}^\top)\mathbf{p}_2\end{bmatrix}\right) = 0, \tag{22}$$

which is possible only when $\mathbf{p}_1, \mathbf{p}_2, \mathbf{n}_1$, and $\mathbf{n}_2$ are co-planar. □

There exist degenerate cases when the conditions of Proposition 2 do not hold. One example is when mirrors are perfectly symmetric, as in Fig. 7, and consequently a reflected ray is on the same plane as the epipolar plane of the true label. However, placing the projector-camera pair in an asymmetric configuration relative to the mirrors avoids such degeneracies. Note that Lanman et al. [2009] used a system that is in a perfectly symmetric configuration by design. Their focus was on mitigating interference, with labeling done manually.

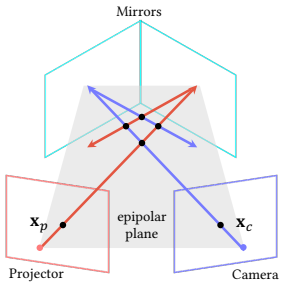
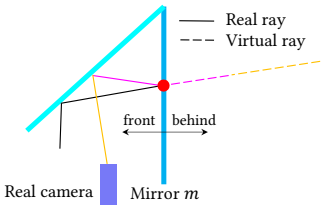


Fig. 7. **Degenerate configuration.** When the mirrors and the rays from the projector and camera are perfectly symmetric, the reflected rays can lie on the epipolar plane of corresponding projector-camera pixels.

From Propositions 1 and 2, we know that minimizing (12) provides the true labels for projector and camera up to mirroring by a shared third label. We will resolve the remaining ambiguity with Proposition 3 below. We first introduce the following observation.

Observation. For a physically feasible traced ray involving mirror reflections, the virtual ray that is unfolded from the real ray before hitting mirror m is always behind the mirror m. The inset shows an example: the virtual rays (dotted line segments) unfolded from the real rays (solid line segments) are always behind the mirror m. This is true because the mirror reflection transforms a real or virtual point in front of the mirror to a virtual point behind the mirror.



With this observation in mind, we can now resolve the ambiguity due to identically mirrored labels.

Proposition 3 (Triangulation from identically mirrored labels.). *The triangulated point from identically mirrored labels is always outside the mirror system.*

Proof. We denote by $\mathbf{X}^*$ the true 3D point triangulated from $\mathbf{x}_p$ and $\mathbf{x}_{c_i}$ given the true labels ℓ_p^* and $\ell_{c_i}^*$. When we have incorrect labels $M(\ell_p^*, \ell')$ and $M(\ell_{c_i}^*, \ell')$ that are identically mirrored with a label $\ell' = (l_1', l_2', ..., l_{K'}')$, the corresponding virtual projectors and cameras are transformed by the label ℓ' relative to the virtual projectors and cameras of the true labels. Therefore, triangulation reconstructs the transformed point $\mathbf{X}_t = \mathbf{D}(\ell')^{-1}\mathbf{X}^*$. We will show that $\mathbf{X}_t$ is always outside the mirror system, and specifically, behind the mirror $l_{K'}'$, corresponding to the last element of ℓ'.

We can prove this directly using the above observation because: 1) the empty label is from the object-free traced ray, which is physically feasible; 2) $\mathbf{X}^*$ is the true 3D point, which is on the real ray before hitting mirror $l_{K'}'$; and 3) $\mathbf{X}_t = \mathbf{D}(l_{K'}')^{-1} \cdots \mathbf{D}(l_1')^{-1}\mathbf{X}^*$ corresponds to a point on the unfolded virtual ray from mirror $l_{K'}'$. Thus, $\mathbf{X}_t$ is behind the mirror $l_{K'}'$, and therefore outside the mirror system. □

In summary, Propositions 1-3 establish that we can determine the labels associated with a projector pixel and its corresponding camera pixels by searching over the empty labels associated with each pixel for the labels that have zero (or smallest) total virtual epipolar distance, and produce a triangulated point inside the mirror system. In the absence of noise, this procedure provably produces correct labels.

Remark. We note that, even though we only discuss *adding* labels to the true label, our derivation automatically covers *removing* labels as well. This is a consequence of the fact that the matrix $\mathbf{D}_m$, which describes the mirror transformation, is involutory. Hence, adding the trailing end of a mirror sequence in reverse is equivalent to deleting the trailing end from the sequence.

4.4 Comparison to Other Labeling Methods

Visual hull. Reshetouski et al. [2011] propose solving the labeling problem using the visual hull, approximated through space carving from background pixels. For objects with simple shapes that are predominantly convex, the labeling from a crude visual hull is often a good approximation to the true labeling. However, for objects that have complex geometry and self-occlusions, the visual hull invariably fails to capture concavities, especially when there are not enough background pixels. Fig. 8 shows an example of such a failure case. (For the visualization of labels, we sorted the mirror sequence in ascending order and used the "cool" MATLAB colormap—magenta-to-cyan linearly.) For such objects, the resulting labeling can significantly deviate from the correct labeling. The labeling accuracy for projector and camera is 76.13 % and 83.47 % for the visual hull method, and 99.96 % and 99.99 % for ours. The visual hull results also depend on the quality of background segmentation and the initial shape for carving. In the simulation for Fig. 8, we used the ground-truth background segmentation and set the initial shape to be a cube that is 20% larger than the ground-truth shape in each axis. Ihrke et al. [2012] used a structured light system with a kaleidoscope as we do, but relied on the inaccurate labeling from the visual hull, resulting in artifacts as we discuss in Section 7.

Pulsed ToF. Xu et al. [2018] combined a kaleidoscope with a pulsed ToF camera, with the source and detector collocated. The ToF measurement provides a simple solution to the labeling problem: As the source and detector are collocated, simply ray tracing from the detector pixel for a distance equal to half the measured ToF provides the location of the 3D point. However, this technique requires a high-cost pulsed ToF camera for precise ToF measurements.

5 SURFACE RECONSTRUCTION

In the previous section, we have shown how to establish *labeled* correspondences $\mathbf{x}_p \leftrightarrow \{\mathbf{x}_{c_i}\}$ between a projector pixel $\mathbf{x}_p$ and multiple camera pixels $\{\mathbf{x}_{c_i}\}$. We now explain how to recover a 3D point $\mathbf{Q}$ from a labeled correspondence, as well as how to reconstruct the surface of the scanned object from multiple such 3D points.

Multi-view triangulation. A naive approach for reconstructing 3D points $\mathbf{Q}$ from a labeled correspondence $\mathbf{x}_p \leftrightarrow \{\mathbf{x}_{c_i}\}$ would be to apply the classical two-view triangulation procedure [Hartley and Zisserman 2004] to each projector-camera pixel pair $\mathbf{x}_p \leftrightarrow \mathbf{x}_{c_i}$ separately. However, this approach does not take into account the information that each of these projector-camera pixel pairs is an observation of the same underlying 3D point $\mathbf{Q}$. These multiple observations of the same point correspond to different viewpoints and baselines, and thus taking them into account jointly can greatly improve the robustness of the reconstruction of the unique 3D point $\mathbf{Q}$. The naive approach produces multiple perturbed versions of this point, resulting in a noisy and redundant point cloud that impedes subsequent surface reconstruction procedures.

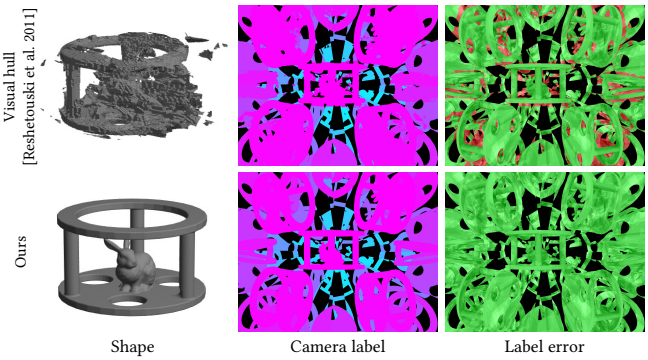


Fig. 8. **Comparison to visual hull.** Reshetouski et al. [2011] solved the labeling problem using the visual hull. Their approach fails when there are insufficient background pixels, or for non-convex objects. Green and red colors in "label error" indicate correct and incorrect labels, respectively.

We adopt a *multi-view triangulation* approach, which uses all the geometric information from the one-to-multiple labeled correspondence $\mathbf{x}_p \leftrightarrow \{\mathbf{x}_{c_i}\}$, to reconstruct a single 3D point $\mathbf{Q}$. We obtain this point by solving a linear least-squares problem:

$$\mathbf{Q} = \underset{\mathbf{Q}}{\operatorname{argmin}} \sum_i h_i^2 \tag{23}$$

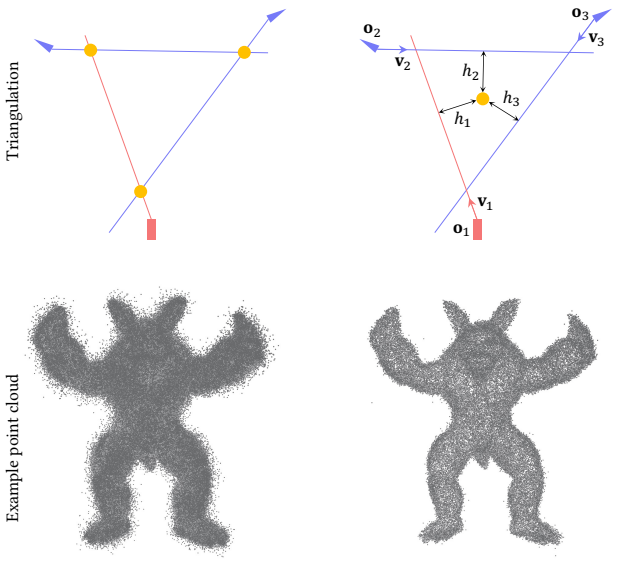
$$= \underset{\mathbf{Q}}{\operatorname{argmin}} \sum_i \| \left(\mathbf{I} - \mathbf{v}_i \mathbf{v}_i{}^\top\right)(\mathbf{Q} - \mathbf{o}_i) \|^2 \tag{24}$$

$$= \underset{\mathbf{Q}}{\operatorname{argmin}}\, \mathbf{Q}^\top \mathbf{A} \mathbf{Q} - 2\mathbf{b}^\top \mathbf{Q} + c \tag{25}$$

$$= \mathbf{A}^{-1}\mathbf{b}, \tag{26}$$

where h_i is the distance from the 3D point $\mathbf{Q}$ to each ray, $\mathbf{o}_i$ is the ray origin, $\mathbf{v}_i$ is the ray direction, and we use $\mathbf{A} \equiv \sum_i \left(\mathbf{I} - \mathbf{v}_i \mathbf{v}_i{}^\top\right)$, and $\mathbf{b} \equiv \sum_i \mathbf{o}_i^\top \left(\mathbf{I} - \mathbf{v}_i \mathbf{v}_i{}^\top\right) \mathbf{o}_i$. Fig. 9 compares point clouds produced using the naive two-view and our multi-view triangulation procedures. For this comparison, we use simulated data where we perturb measurements of camera pixel locations with Gaussian noise of variance 5 pixels. We observe that the point cloud from multi-view triangulation is less noisy than that from two-view triangulation. We empirically found this to be the case across all our experiments. Therefore, we use multi-view triangulation throughout the paper.

Outlier rejection. The multi-view triangulation estimate of (26) is susceptible to outlier rays due to incorrect pixel detections (e.g., pixels illuminated due to direct illumination of the camera from a virtual projector, or indirect illumination effects). Such outliers can cause severe errors in the estimation of the unknown 3D point



(a) Two-view triangulation (b) Multi-view triangulation

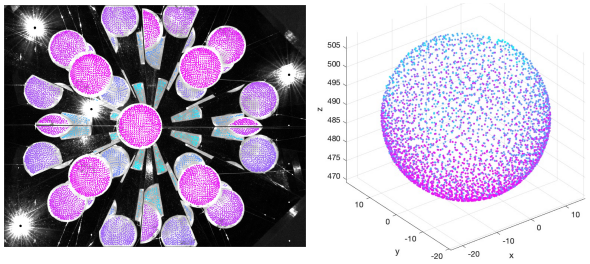
Fig. 9. **Triangulation.** Given the correspondence between a projector pixel and multiple camera pixels, we can perform either (a) two-view triangulation for all pairs in the correspondence; or (b) multi-view triangulation. Multi-view triangulation reconstructs a single 3D point that is closest to all the rays, making it more robust than two-view triangulation. Multi-view triangulation produces less noisy results than two-view triangulation when there is noise in the measurements.

Table 2. **Components used in our hardware prototype.**

Description	Company	Model name
laser projector	Sony	MP-CL1A
camera	FLIR	GS3-U3-91S6M
camera lens	Nikon	AF Nikkor 24mm f/2.8D
mirror	Edmund Optics	46-656 (custom)

Q. We address this issue using RANSAC [Fischler and Bolles 1981]: We repeatedly perform two-view triangulation between projector-camera pixel pairs $\mathbf{x}_p \leftrightarrow \mathbf{x}_{c_i}$ randomly selected from the labeled correspondence $\mathbf{x}_p \leftrightarrow \{\mathbf{x}_{c_i}\}$. For each reconstructed 3D point, we choose as inliers the pixels in $\{\mathbf{x}_{c_i}\}$ whose corresponding rays are close enough (0.5 mm) to the reconstructed point. Finally, we perform multi-view triangulation between $\mathbf{x}_p$ and the largest inlier set to get a robust estimate of the 3D point.

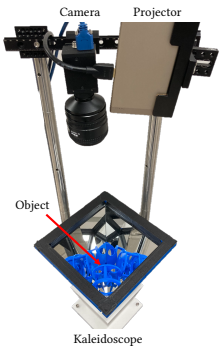
Surface reconstruction. Our triangulation procedure produces a 3D point for each labeled correspondence $\mathbf{x}_p \leftrightarrow \{\mathbf{x}_{c_i}\}$. The last step in our reconstruction pipeline is to use the resulting 3D point cloud to reconstruct the object surface. For this, we first compute PCA normals [Hoppe et al. 1992] for each point in the point cloud. We then reconstruct a mesh representation of the scanned object by using screened Poisson surface reconstruction [Kazhdan and Hoppe 2013], which estimates an implicit surface from the oriented point cloud, and extracts an isosurface.



(a) Manual label (from initial sphere fitting) (b) Bundle adjustment results

Fig. 10. **Calibration.** We calibrate the projector, camera, and mirrors using a reference spherical object of known diameter. We manually label pixels using an initial sphere fitting result, and perform bundle adjustment to minimize triangulation error, sphere fitting error, and reprojection error.

6 IMPLEMENTATION



We have developed a hardware prototype for a kaleidoscopic structured light system, comprising a laser projector, a monochrome CCD camera, and a kaleidoscope. The inset shows a photograph of our prototype. For the kaleidoscope, we use four planar metal-coated mirrors (surface flatness $4-6\lambda$, dimensions 200 mm × 307 mm) that we cut to be shaped as isosceles triangles. Table 2 lists the exact parts used in our prototype. In the rest of this section, we describe how we calibrate our system, and how we accelerate scanning.

Calibration. We calibrate our projector-camera pair using the algorithms of Zhang et al. [2000] and Moreno et al. [2012]. We calibrate the kaleidoscope using the algorithm of Takahashi et al. [2017; 2021], which estimates the location and pose of the mirrors relative to the camera from correspondences of a single 3D point.

To improve upon this initial calibration, we use a bundle adjustment procedure inspired by Xu et al. [2018]. As we show in Fig. 10, we optimize the parameters of the projector, camera, and mirrors, based on scanning results for a reference object (sphere of diameter 40 mm). We first sparsely scan the reference sphere for a few pixels that can be easily labeled manually (e.g., direct and one-bounce), and fit a sphere to the reconstructed point cloud. With this initial sphere fitting result, we can label every pixel using ray tracing, and completely reconstruct the object. Then, we update the extrinsic parameters of the projector and mirrors relative to the camera by minimizing an objective combining triangulation error (distance of a reconstructed point from its corresponding backprojected rays), reprojection error (distance of the projection of a reconstructed point from its corresponding pixel locations), and sphere fitting error of the reconstructed point cloud. After bundle adjustment, we achieve a root-mean-square triangulation error of 33 µm, reprojection error of 1.3 pixels, and sphere fitting error of 361 µm.

Parallel scanning. Up to this point, we have assumed that scanning works by sequentially illuminating all projector pixels, one

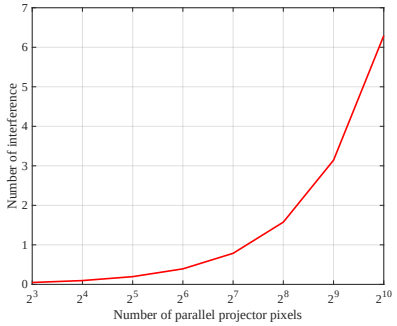


Fig. 11. **Interference in parallel scanning of a spherical object.** The graph quantifies the trade-off between interference and acquisition speed, by plotting the number of points interfered by mirror reflection against the number of simultaneously illuminated projector pixels.

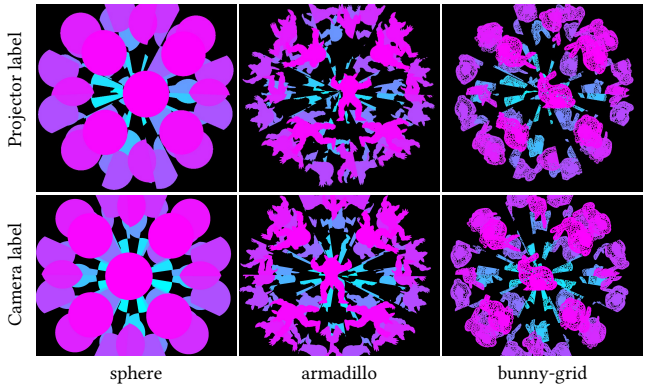


Fig. 12. **Labeling accuracy for synthetic data.** We visualize the labeling results under Gaussian noise for several simulated objects.

Table 3. **Labeling accuracy statistics.** We report labeling accuracy metrics, with and without adding Gaussian noise ($\sigma = 5$ pixels) to camera pixel measurements. The labeling is robust to measurement noise.

Labeling accuracy	sphere	armadillo	bunny-grid
Projector (w/o noise)	100.00%	100.00%	100.00%
Projector (Gaussian noise)	99.69%	99.69%	99.43%
Camera (w/o noise)	100.00%	99.99%	100.00%
Camera (Gaussian noise)	99.99%	99.99%	99.98%

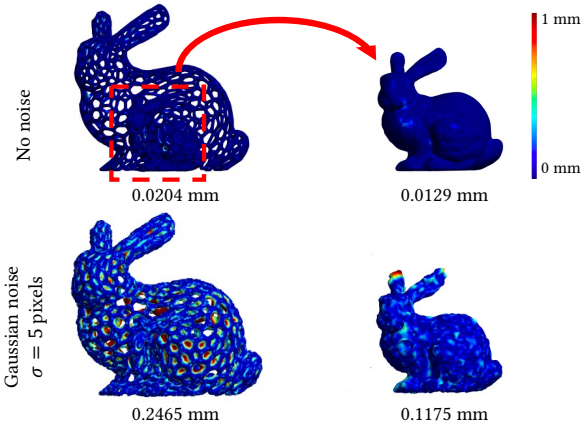


Fig. 13. **Reconstruction accuracy for synthetic data.** We simulate scanning of an object with diameter 60 mm using the same imaging system as our prototype, with and without noise. Our system makes it possible to accurately reconstruct the severely occluded inner bunny.

at a time. However, sequential scanning makes acquisition times impractically long. We now describe a parallel scanning technique for faster acquisition.

Similar to traditional structured light techniques, our parallel scanning technique operates by simultaneously illuminating multiple projector pixels with a temporal code, and decoding the measurements made at a pixel to obtain projector-camera correspondences. However, a key difference in the kaleidoscopic setting is the complexity of the epipolar geometry (and thus, interference patterns) produced by the multiple mirrors, which makes traditional column-scanning schemes inapplicable. Instead, we *randomly* sample groups of $2^8 - 1$ projector pixels, encode each pixel with an 8-bit binary code, and project the sequence of binary images (and their inverses for improved robustness) that correspond to each bit of the binary code. Then, we obtain the camera pixels corresponding to each projector pixel by decoding the binary code from the captured image.

This parallel scanning technique accelerates acquisition, but also exacerbates interference: multiple projector pixels may illuminate the same 3D point, which could result in erroneous decoding. The likelihood of this happening increases as we increase the number of projector pixels we simultaneously illuminate, which creates a trade-off between acquisition acceleration and interference.

To empirically quantify this trade-off, we simulated parallel scanning of a spherical object. Fig. 11 plots the number of points having interference in each scan, as a function of the number of simultaneously illuminated projector pixels. Based on this plot, we chose to illuminate $2^8 - 1$ projector pixels, corresponding to an average of one pixel with interference in each scan. This results in few decoding errors, which are easily handled during triangulation via RANSAC.

7 RESULTS

We evaluate our method using simulated and real experiments. Our code and data are available on the project page [Ahn et al. 2021].

7.1 Simulated Experiments

We use a ray tracing implementation (customized to handle multiple specular-specular reflections) to simulate measurements from our kaleidoscopic structured light system. We use these simulated measurements to evaluate the results of our labeling and shape reconstruction procedures against known ground truth. The diameter of our simulated objects is 60 mm.

Labeling accuracy. Fig. 12 and Table 3 show simulated labeling results for three synthetic objects, with and without Gaussian noise in

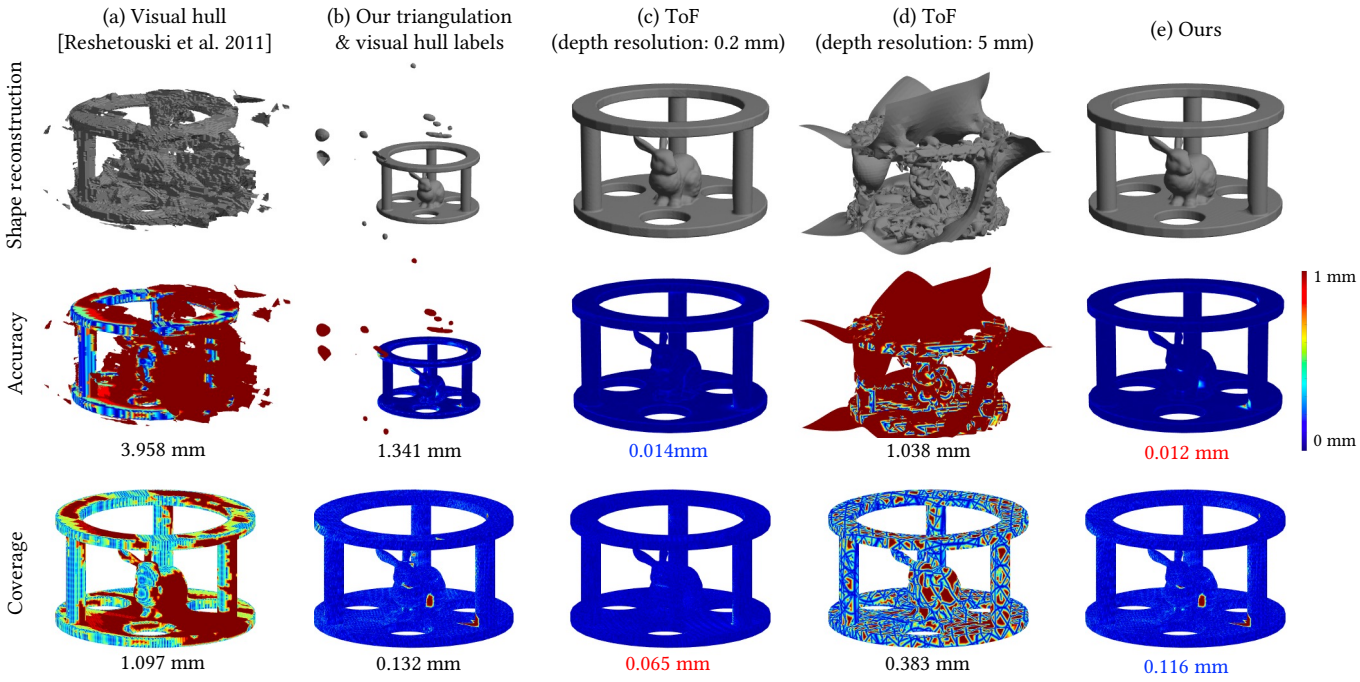


Fig. 14. **Simulated comparison of kaleidoscopic imaging methods.** We use simulation to quantitatively evaluate the performance of various kaleidoscopic imaging methods. Our method provides the best reconstruction accuracy.

the camera pixel measurements ($\sigma = 5$ pixels). Our labeling provides almost 100% accuracy in all cases.

Reconstruction accuracy. Fig. 13 shows simulated shape reconstruction results, and average distance between reconstruction and ground truth. We note that the multiple virtual views of our system make it possible to reconstruct the severely-occluded inner bunny.

Comparisons with other kaleidoscopic imaging methods. Fig. 14 shows simulated comparisons to other kaleidoscopic imaging methods. We used pixel binning and depth binning to simulate the finite resolution of the sensor used in each method. We compare the performance of different approaches using two metrics: The first metric is *accuracy*, which we define as the average distance of the vertices of the reconstructed mesh from the ground-truth mesh; this metric quantifies how the reconstruction is to the ground truth. The second metric is *coverage*, which we define as the average distance of the vertices of the ground-truth mesh from the nearest point of the reconstructed point cloud; this metric quantifies how well the reconstruction covers each part of the ground-truth shape.

- Fig. 14(a) shows results from the visual hull technique proposed by Reshetouski et al. [2011]. We observe that this technique does not recover the inner bunny, because of the limited background area available for space carving.
- Fig. 14(b) shows results from a variation of our technique inspired by Ihrke et al. [2012], where we perform multi-view triangulation with RANSAC, but using labels from the visual hull (Fig. 8). We observe that, even though RANSAC mitigates the effect of incorrect labels, there are still some artifacts remaining.
- Fig. 14(c) and (d) show results from the ToF-based technique of Xu et al. [2018]. We set the spatial resolution of the ToF camera to be the same as that of our camera, and simulate two different ToF depth resolutions: Fig. 14(c) uses a depth resolution of 0.2 mm, corresponding to the calibration error reported by Xu et al. [2018] using a costly high-end lidar (Leica ScanStation P40 3D Laser Scanner). Fig 14(d) uses a depth resolution of 5 mm, corresponding to a low-end lidar of cost comparable to our setup (Intel Realsense LiDAR Camera L515). We observe that lowering the resolution results in a severe increase in noise in the reconstructed shape.
- Fig. 14(e) shows results from our method. We observe that, compared to the visual hull technique (a), our use of triangulation enables reconstructing the concave and occluded parts of the shape. Additionally, using accurate labels from our labeling technique reduces artifacts and improves accuracy by two orders of magnitude compared to using visual hull labels (b). Our technique has a similar accuracy improvement when compared to ToF using a lidar of cost comparable to our setup (d). Lastly, our technique achieves comparable accuracy with the ToF setup of Xu et al. [2018], while using a much more affordable imaging setup.

7.2 Real Experiments

Scanned objects. Fig. 15 shows reconstructions of a variety of real objects obtained using the kaleidoscopic structured light prototype of Sec. 6. For each object, we show a kaleidoscopic image under uniform projector illumination, camera labels, and reconstructed mesh surface. Our setup allows scanning objects of size up to about 10 cm (e.g., the skull has dimensions $5 \times 10 \times 7.5$ cm^3). To visualize

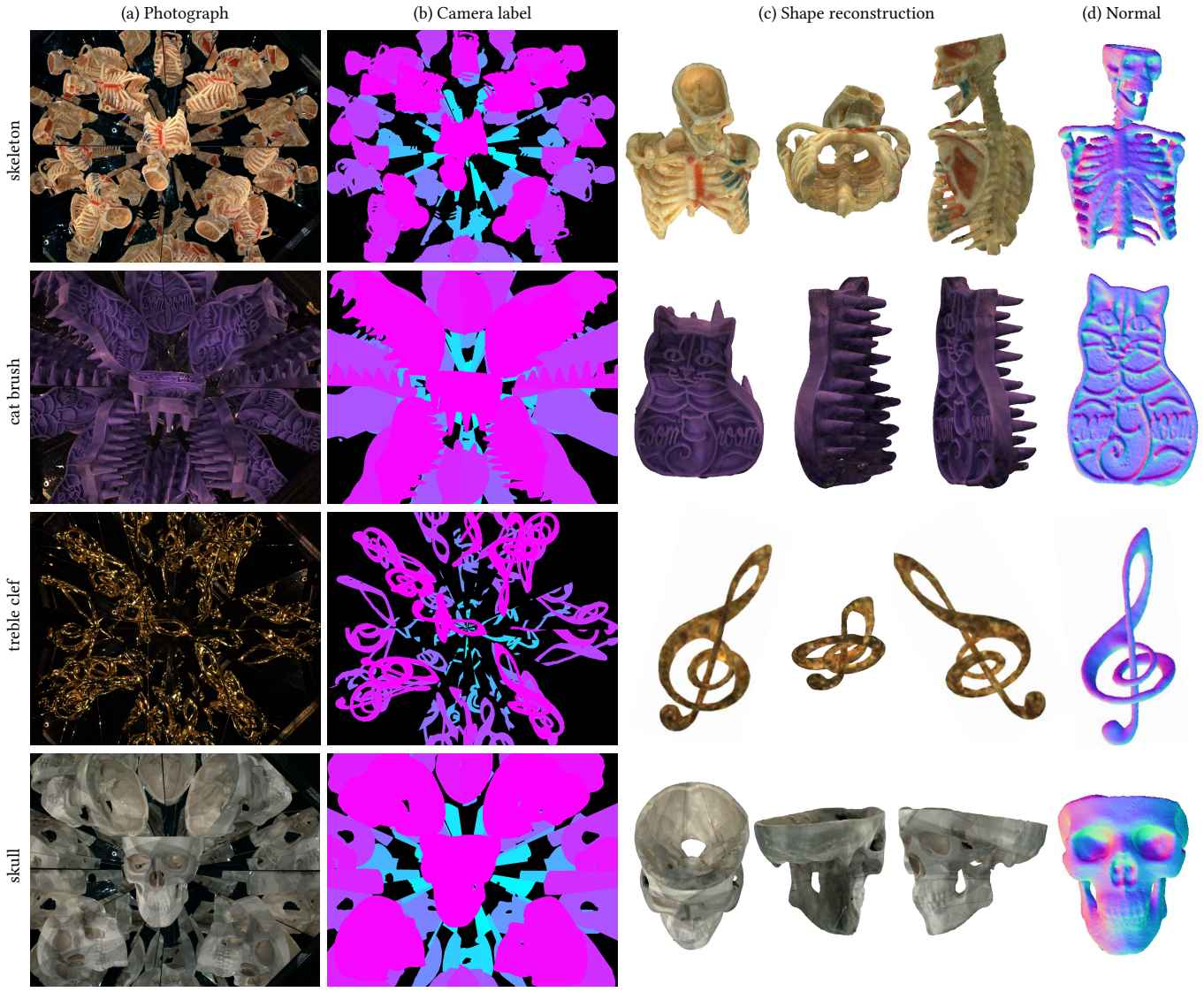


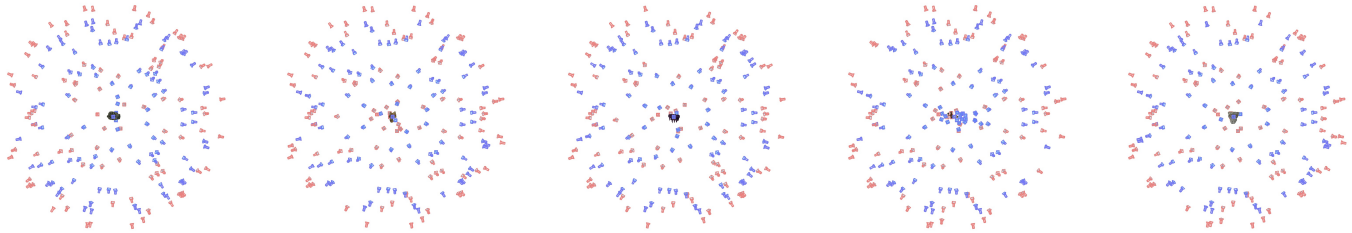
Fig. 15. **Real object scans from our prototype.** Reconstructed point clouds and full surround videos are available on the project webpage [Ahn et al. 2021].

the appearance of the reconstructed surface, we use a simple texture mapping procedure by projecting each vertex to all visible (virtual) cameras, and computing the average intensity of the corresponding pixel values. As our setup uses a monochrome camera, we obtain per-pixel color by projecting color channels sequentially from our RGB projector. We additionally visualize PCA vertex normals, to help better assess the quality of the reconstructed mesh. Fig. 1 shows an additional scanned object. We observe that our kaleidoscopic structured light system produces high-quality reconstructions for a variety of objects with complex visibility and reflectance properties.

Visibility. Fig. 16 and Table 4 report the effective average number of projector views, camera views, and unique projector-camera pairs per mesh vertex, for the scanned objects of Fig. 1 and Fig. 15. These numbers are strongly affected by the size and location of the object inside the kaleidoscope. For example, the skull has an average of 22.2 projector-camera pairs per vertex, whereas the smaller treble clef has an average of 96.5 pairs per vertex.

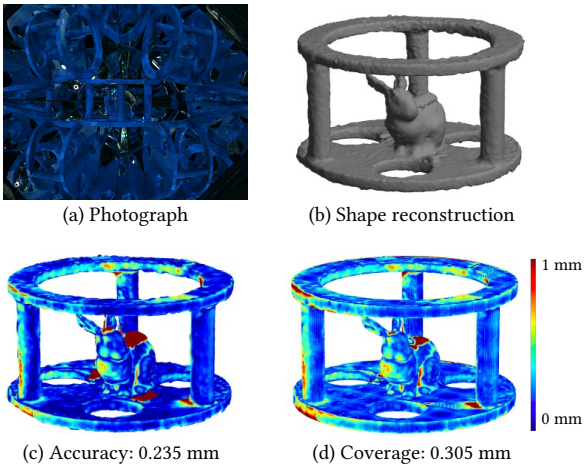
Table 4. **Effective number of per-vertex projector and camera views.** The number is generally smaller for larger objects because of occlusion.

#views	elephant	skeleton	cat brush	treble clef	skull
projector	5.2	6.3	4.8	9.1	4.7
camera	5.4	6.9	4.5	10.2	4.3
pair	31.8	51.5	24.3	96.5	22.2



elephant: (88, 93) skeleton: (91, 86) cat: (91, 89) treble clef: (85, 99) skull: (88, 83)

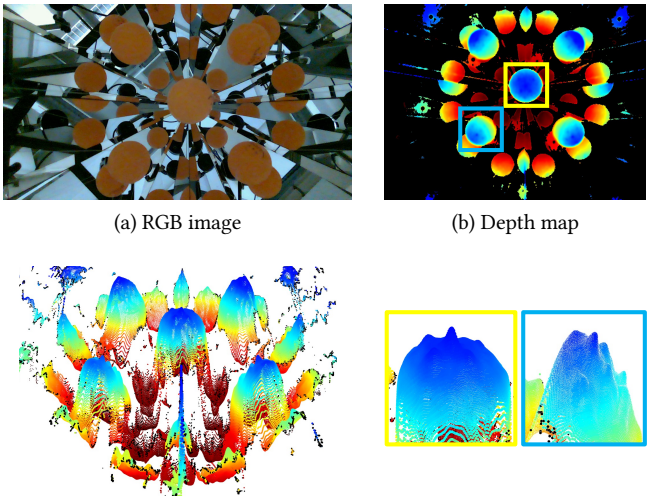
Fig. 16. **Effective projector and camera views for different objects.** We visualize and report the number of the effective projectors and cameras around the object. Overall, there are almost 100 virtual projectors and cameras surrounding each object.



(a) Photograph (b) Shape reconstruction

(c) Accuracy: 0.235 mm (d) Coverage: 0.305 mm

Fig. 17. **Real scan of a 3D-printed mesh.** We 3D-printed and scanned the mesh used in the simulation results of Fig. 14. We report performance metrics comparing the reconstructed and ground-truth meshes.



(a) RGB image (b) Depth map

(c) Point cloud (unfolded) (d) Zoom in (noisy surface)

Fig. 18. **Kaleidoscopic time-of-flight experiment.** Unfolded point cloud obtained from the depth map of a commercial low-resolution lidar. The depth measurements are a lot noisier than those obtained using our kaleidoscopic structured light system for the same object (Fig. 10).

Quantitative evaluation. To quantify the reconstruction accuracy and coverage of our technique, we used our setup to scan a 3D-printed object for which a ground-truth mesh is available. The object had a width of 8 cm, and was 3D-printed at a layer resolution of 0.17 mm. Fig. 17 shows the results. We aligned the reconstructed mesh with the ground-truth one using the iterative closest point algorithm [Besl and McKay 1992]. By comparing the two meshes, we estimated an accuracy of 0.235 mm and coverage of 0.305 mm for our method. By comparing these numbers with those in Fig. 14(e), where we used the same ground-truth mesh for a simulated experiment, we can also quantitatively assess the impact of calibration errors and other hardware imperfections on reconstruction quality.

Comparison with kaleidoscopic ToF system. We performed an experiment where we replaced the projector-camera pair in our prototype with a commercial lidar of cost comparable to our setup (Intel Realsense LiDAR Camera L515, same depth resolution as that used for the simulated experiments of Fig 14(d)). We used this modified system to scan the same spherical calibration object as in Fig. 10. We show the results in Fig. 18. We note that, because of the low depth resolution and noisy ToF measurements of our lidar, we were unable to obtain accurate calibration information for the modified imaging setup using the calibration procedure of Xu et al. [2018]. In turn, the lack of accurate calibration meant we could not produce a meaningful shape reconstruction using their proposed ray folding procedure. We show, instead, our reconstructed *unfolded* point cloud, which we observe to be a lot noisier than the one obtained by our technique (Fig. 10). As Xu et al. [2018] and our simulations in Fig. 14 both have shown, using a high-end lidar can alleviate these issues and achieve the same reconstruction accuracy as our structured light technique, albeit at a much higher hardware cost.

8 DISCUSSION

PCA normals. The combined use of our labeled correspondences and multi-view triangulation with RANSAC robustly reconstructs

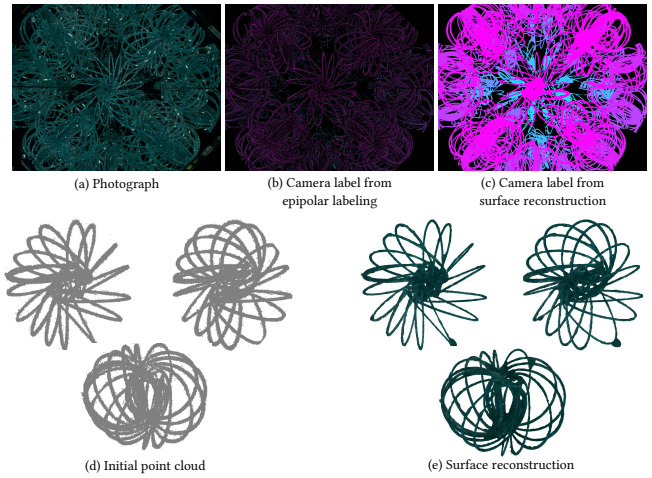


Fig. 19. **Effect of PCA normals.** We show an example of how of failures in normal estimation impact reconstruction quality: Our system estimates an accurate point cloud of a very thin object, for which the PCA normal estimates are inaccurate. As a result, the reconstructed mesh has strong artifacts, especially at the center where different rings intersect.

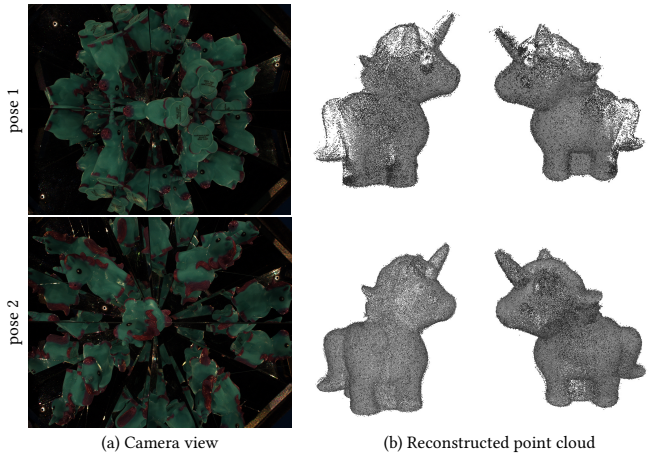


Fig. 20. **Effect of object pose inside the kaleidoscope.** Changing the object pose impacts reconstruction quality. Having enough space between the object and mirrors can reduce occlusion and improve the reconstruction.

accurate point clouds. However, the screened Poisson surface reconstruction [Kazhdan and Hoppe 2013] algorithm we use to create the final mesh reconstructions requires as input *oriented* point clouds. We produce those by assigning PCA normals [Hoppe et al. 1992] to our reconstructed unoriented point clouds. Unfortunately, PCA normals can be inaccurate for shapes with complex topology, resulting in inaccurate meshes. As an example, in Fig. 19 we show scan results for a slinky: The reconstructed point cloud accurately represents the object's interwoven thin parts; however, the reconstructed mesh has strong artifacts near the center, because of the inaccurate PCA normals. Combining our technique with more accurate normal estimation procedures, including ones using shading information, can help improve the accuracy of the final mesh reconstructions.

Pose of the object in the kaleidoscope. Fig. 20 shows results from scans of the same object from two different poses: in pose 1, we placed the object directly on the kaleidoscope, whereas in pose 2, we hung it using strings. In the former case, the parts on the mirrors (e.g., head, tail) are visible from few viewpoints, and thus are not reconstructed. By contrast, in the latter case, these parts are well reconstructed. This example highlights the strong impact object pose can have on the final reconstruction, and suggests object pose optimization as an important future research direction.

Comparison to neural rendering. 3D reconstruction by moving a flash-camera pair around an object, in the style of IDR [Yariv et al. 2020] and NeRF [Mildenhall et al. 2020], has recently seen immense success. This is advantageous over our technique in terms of cost, object size, acquisition time, and overall user convenience. By contrast, our structured light technique can handle textureless objects where passive techniques fail. Our technique can also handle very complex shapes where, due to visibility, flash photography methods can produce poor results, unless the number of views becomes impractically large. In the future, it is possible that using neural rendering algorithms to process measurements from kaleidoscopic structured light setups can lead to scanning technologies that combine the complementary advantages of the two approaches.

9 CONCLUSION

We introduced a full surround 3D imaging technique that combines a projector-camera pair with a kaleidoscope, to produce a virtual multi-view structured light system. We derived an algorithm that uses the epipolar geometry between virtual projectors and virtual cameras to produce provably correct labels for pixels in the kaleidoscopic image, in terms of which virtual projector and virtual camera these pixels correspond to. By combining these labels with multi-view triangulation, we showed that our system can achieve high reconstruction accuracy and full coverage, even when scanning objects with complex geometry and reflectance.

ACKNOWLEDGMENTS

We thank Jinmo Rhee and Vishwanath Saragadam for help with the prototype. This work was supported by the National Science Foundation (NSF) under awards 1652569, 1900849, and 2008464, as well as a Sloan Research Fellowship for Ioannis Gkioulekas.

REFERENCES

Kfir Aberman, Oren Katzir, Qiang Zhou, Zegang Luo, Andrei Sharf, Chen Greif, Baoquan Chen, and Daniel Cohen-Or. 2017. Dip transform for 3D shape reconstruction. *ACM Transactions on Graphics (TOG)* 36, 4 (2017).

Byeongjoo Ahn, Ioannis Gkioulekas, and Aswin C. Sankaranarayanan. 2021. *Project page: Kaleidoscopic structured light.* https://imaging.cs.cmu.edu/kaleidoscopic_structured_light

Shaun Bangay and Judith D Radloff. 2004. Kaleidoscope configurations for reflectance measurement. In *International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa.*

Paul J Besl and Neil D McKay. 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 14, 2 (1992).

David Brewster. 1858. *The kaleidoscope, its history, theory and construction: with its application to the fine and useful arts.* J. Murray.

Yan Cui, Sebastian Schuon, Derek Chan, Sebastian Thrun, and Christian Theobalt. 2010. 3D shape scanning with a time-of-flight camera. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Martin A Fischler and Robert C Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (1981).

Keith Forbes, Fred Nicolls, Gerhard De Jager, and Anthon Voigt. 2006. Shape-from-silhouette with two mirrors and an uncalibrated camera. In *European Conference on Computer Vision (ECCV)*.

Martin Fuchs, Markus Kächele, and Szymon Rusinkiewicz. 2013. Design and fabrication of faceted mirror arrays for light field capture. In *Computer Graphics Forum*, Vol. 32.

Gaurav Garg, Eino-Ville Talvala, Marc Levoy, and Hendrik PA Lensch. 2006. Symmetric Photography: Exploiting data-sparseness in reflectance fields. In *Symposium on Rendering*.

Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics (TOG)* 30, 6 (2011).

Joshua Gluckman and Shree K Nayar. 2001. Catadioptric stereo using planar mirrors. *International Journal of Computer Vision (IJCV)* 44, 1 (2001).

Joshua Gluckman and Shree K Nayar. 2002. Rectified catadioptric stereo sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 24, 2 (2002).

Ardeshir Goshtasby and William A Gruver. 1993. Design of a single-lens stereo camera system. *Pattern Recognition* 26, 6 (1993).

Jefferson Y Han and Ken Perlin. 2003. Measuring bidirectional texture reflectance with a kaleidoscope. 22, 3 (2003).

Richard Hartley and Andrew Zisserman. 2004. *Multiple view geometry in computer vision* (2nd ed.). Cambridge University Press.

Michael Holroyd, Jason Lawrence, and Todd Zickler. 2010. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Transactions on Graphics (TOG)* 29, 4 (2010).

Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. 1992. Surface Reconstruction from Unorganized Points. *Computer Graphics (SIGGRAPH'92 proceedings)* 26, 2 (1992).

Bo Hu, Christopher Brown, and Randal Nelson. 2005. *Multiple-view 3-D reconstruction using a mirror.* Technical Report. University of Rochester, Department of Computer Science.

Po-Hao Huang and Shang-Hon Lai. 2006. Contour-based structure from reflection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ivo Ihrke, Ilya Reshetouski, Alkhazur Manakov, Art Tevs, Michael Wand, and Hans-Peter Seidel. 2012. A kaleidoscopic approach to surround geometry and reflectance acquisition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.

Hanbyul Joo, Tomas Simon, Xulong Li, Hao Liu, Lei Tan, Lin Gui, Sean Banerjee, Timothy Godisart, Bart Nabbe, Iain Matthews, et al. 2017. Panoptic studio: A massively multiview system for social interaction capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 41, 1 (2017).

Kaizhang Kang, Cihui Xie, Chengan He, Mingqi Yi, Minyi Gu, Zimin Chen, Kun Zhou, and Hongzhi Wu. 2019. Learning efficient illumination multiplexing for joint capture of reflectance and shape. *ACM Transactions on Graphics (TOG)* 38, 6 (2019).

Michael Kazhdan and Hugues Hoppe. 2013. Screened poisson surface reconstruction. *ACM Transactions on Graphics (TOG)* 32, 3 (2013).

Kalin Kolev, Petri Tanskanen, Pablo Speciale, and Marc Pollefeys. 2014. Turning mobile phones into 3D scanners. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kiriakos N Kutulakos and Steven M Seitz. 2000. A theory of shape by space carving. *International Journal of Computer Vision (IJCV)* 38, 3 (2000).

Douglas Lanman, Daniel Crispell, and Gabriel Taubin. 2009. Surround structured lighting: 3-D scanning with orthographic illumination. *Computer Vision and Image Understanding (CVIU)* 113, 11 (2009).

Daniel Lichy, Jiaye Wu, Soumyadip Sengupta, and David W Jacobs. 2021. Shape and material capture at home. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2020. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*.

Hiroshi Mitsumoto, Shinichi Tamura, Kozo Okazaki, Naoki Kajimi, and Yutaka Fukui. 1992. 3-D reconstruction using mirror images based on a plane symmetry recovering method. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 14, 09 (1992).

Daniel Moreno and Gabriel Taubin. 2012. Simple, accurate, and robust projector-camera calibration. In *IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*.

David W Murray. 1995. Recovering range using virtual multicamera stereo. *Computer Vision and Image Understanding (CVIU)* 61, 2 (1995).

Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H Kim. 2018. Practical svbrdf acquisition of 3d objects with unstructured flash photography. *ACM Transactions on Graphics (TOG)* 37, 6 (2018).

Sameer A Nene and Shree K Nayar. 1998. Stereo with mirrors. In *IEEE International Conference on Computer Vision (ICCV)*.

Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. 2011. KinectFusion: Real-time dense surface mapping and tracking. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*.

NextEngine. 2000. . http://www.nextengine.com [Accessed Sep. 9, 2021].

Shohei Nobuhara, Takashi Kashino, Takashi Matsuyama, Kouta Takeuchi, and Kensaku Fujii. 2016. A single-shot multi-path interference resolution for mirror-based full 3D shape measurement with a correlation-based ToF camera. In *International Conference on 3D Vision (3DV)*.

Peter Ondrúška, Pushmeet Kohli, and Shahram Izadi. 2015. Mobilefusion: Real-time volumetric surface reconstruction and dense tracking on mobile phones. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 21, 11 (2015).

Jaesik Park, Sudipta N Sinha, Yasuyuki Matsushita, Yu-Wing Tai, and In So Kweon. 2016. Robust multiview photometric stereo using planar mesh parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 39, 8 (2016).

Ilya Reshetouski, Alkhazur Manakov, Hans-Peter Seidel, and Ivo Ihrke. 2011. Three-dimensional kaleidoscopic imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Christopher Schwartz, Ralf Sarlette, Michael Weinmann, and Reinhard Klein. 2013. DOME II: a parallelized BTF acquisition system. In *Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition*.

Yuichi Taguchi, Amit Agrawal, Srikumar Ramalingam, and Ashok Veeraraghavan. 2010a. Axial light field for curved mirrors: Reflect your perspective, widen your view. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yuichi Taguchi, Amit Agrawal, Ashok Veeraraghavan, Srikumar Ramalingam, and Ramesh Raskar. 2010b. Axial-cones: Modeling spherical catadioptric cameras for wide-angle light field rendering. *ACM Transactions on Graphics (TOG)* 29, 6 (2010).

Tomu Tahara, Ryo Kawahara, Shohei Nobuhara, and Takashi Matsuyama. 2015. Interference-free epipole-centered structured light pattern for mirror-based multi-view active stereo. In *International Conference on 3D Vision (3DV)*.

Kosuke Takahashi, Akihiro Miyata, Shohei Nobuhara, and Takashi Matsuyama. 2017. A linear extrinsic calibration of kaleidoscopic imaging system from single 3d point. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kosuke Takahashi and Shohei Nobuhara. 2021. Structure of multiple mirror system from kaleidoscopic projections of single 3d point. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2021).

Hongzhi Wu, Zhaotian Wang, and Kun Zhou. 2015. Simultaneous localization and appearance estimation with a consumer RGB-D camera. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 22, 8 (2015).

Hongzhi Wu and Kun Zhou. 2015. AppFusion: Interactive appearance acquisition using a kinect sensor. *Computer Graphics Forum* 34, 6 (2015).

Rui Xia, Yue Dong, Pieter Peers, and Xin Tong. 2016. Recovering shape and spatially-varying surface reflectance under unknown illumination. *ACM Transactions on Graphics (TOG)* 35, 6 (2016).

Ruilin Xu, Mohit Gupta, and Shree K Nayar. 2018. Trapping light for time of flight. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. 2020. Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance. In *Neural Information Processing Systems (NeurIPS)*.

Xianghua Ying, Kun Peng, Ren Ren, and Hongbin Zha. 2010. Geometric properties of multiple reflections in catadioptric camera with two planar mirrors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhengyou Zhang. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 22, 11 (2000).

Zhenglong Zhou, Zhe Wu, and Ping Tan. 2013. Multi-view photometric stereo with spatially varying isotropic materials. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.