

# Multi-view Reconstruction of Highly Specular Surfaces in Uncontrolled Environments

Clément Godard\*, Peter Hedman\*, Wenbin Li and Gabriel J. Brostow  
University College London

{c.godard, p.hedman, w.li, g.brostow}@cs.ucl.ac.uk

## Abstract

Reconstructing the surface of highly specular objects is a challenging task. The shapes of diffuse and rough specular objects can be captured in an uncontrolled setting using consumer equipment. In contrast, highly specular objects have previously deterred capture in uncontrolled environments and have only been reconstructed using tailor-made hardware. We propose a method to reconstruct such objects in uncontrolled environments using only commodity hardware. As input, our method expects multi-view photographs of the specular object, its silhouettes and an environment map of its surroundings.

We compare the reflected colors in the photographs with the ones in the environment to form probability distributions over the surface normals. As the effect of inter-reflections cannot be ignored for highly specular objects, we explicitly model them when forming the probability distributions. We recover the shape of the object in an iterative process where we alternate between estimating normals and updating the shape of the object to better explain these normals.

We run experiments on both synthetic and real-world data, that show our method is robust and produces accurate reconstructions with as few as 25 input photographs.

## 1. Introduction

3D reconstruction of both scenes and objects has become accessible through photogrammetry, depth cameras and affordable laser scanners. These technologies are designed with the assumption that the objects are covered with diffuse materials, whose appearance does not change with the viewpoint. Consequently, objects made of highly specular, mirror-like, materials cannot be reconstructed using those methods unless they have been covered with a diffuse coating. To date, such objects can only be reconstructed in controlled environments, using expensive, tailor-made hardware such as motor controlled illumination [26, 5] or monitors displaying patterns [42].

We propose a method that accurately reconstructs mirror reflective objects in a casual setting using a consumer camera.



Figure 1. **Our method.** From left to right: The initialization from the image silhouettes, our reconstruction, a synthetic rendering of our reconstruction and a real photograph of the object in the same environment.

Provided that the object is sufficiently far away from its surroundings, our method is able to reconstruct the surface of the object based only on a panoramic image of the environment and around 30 calibrated photographs viewing the object from different directions as well as their silhouettes. This enables cost-effective in-the-field reconstruction of highly specular objects, such as historical artifacts that cannot be moved, let alone sprayed with diffuse paint. The comparatively small number of photographs required by our method could also facilitate fast automated quality assessment of metallic mechanical parts.

Similarly to [28], we start from the visual hull of the object’s shape, and build a probability distribution of the surface normals based on correspondences between the colors reflected by the object and those in the environment. We use these distributions to form estimates of the surface normals, which are used to iteratively reconstruct the object and its details. However, we focus on highly specular objects while they target objects with materials containing either a strong diffuse component and/or rough specularities. As such, we cannot rely on the roughness of the material to guide the optimization. Indeed the appearance of a surface with a rough material varies smoothly as its shape changes, this is not the case for highly specular materials.

We present four main contributions:

- A Bayesian framework for estimating surface normals that is robust to outliers by modeling the likelihood of an observed color as a multivariate t-distribution.
- A method to incorporate inter-reflections into the probabilistic

\*Joint first authors.

- estimate of the surface normals by explicitly modeling them.
- A method to estimate frontier points based only on the visual hull and its associated cameras.
  - A publicly available dataset with both real and synthetic specular objects, captured in uncontrolled environments, and with associated ground truth geometry.

We also perform an analysis of the challenges associated with reconstructing highly specular objects in this setting. Namely, we investigate the impact of the proximity of the environments as well as their variety, the handling of inter-reflections and the sampling rate of the surface normals.

## 2. Related work

Our work fits within the larger problem of specular surface reconstruction from images and is therefore related to a number of techniques such as specular flow, shape from distortion and controlled lighting. Refer to Ihrke *et al.* [14] for a general survey.

### 2.1. Specular surface reconstruction

**Specular Flow.** In specular flow, surfaces are derived by tracking reflections (or virtual features [26]) of changing illumination on them. Roth and Black [30] apply diffuse flow to reconstruct a reflective surface containing diffuse markings. Adato *et al.* [2] achieve simple surface reconstruction on real-world data using a specular flow formulation. More recently, Sankaranarayanan *et al.* [32] successfully detect parabolic points using image invariants on reflective surfaces. To refine the specular flow approaches, Canas *et al.* propose a linear formulation to simplify the optimization. They also attempt to recover the surface using a single image [3, 1] and further reduce the number of flows necessary for certain shapes [40].

These methods do not require prior knowledge of the environment but are sensitive to textureless environments (as they rely on specular flow computation) assume a smooth surface without occlusions and do not handle inter-reflections. Furthermore, the capture of real-world objects requires a complicated setup where both the camera and the surface are attached to a structure that rotates by a known motion to generate the needed flow. Our method works on arbitrary closed surfaces, does not rely on any custom built hardware and is easily reproducible.

**Shape From Distortion.** Shape from distortion relies on the deformation of a given pattern reflected by the surface of the object. Bonfort and Sturm [7] use the reflections of a customized pattern to perform voxel carving of the reflective surface. Savarese *et al.* [33, 34] recover a sparse point set and their normals using correspondences between a known checkerboard and its reflection. They also study how humans perceive specular reflections on different surfaces [35]. Tappen [37] studies the recovery of surface normals for smooth, reflective heightfield surfaces from a single photograph in an unknown environment. Balzer *et al.* [5] reconstruct specular

surfaces using the reflections of a known pattern displayed at the end of a controllable robot arm. Similarly Weinmann *et al.* [42] successfully reconstruct mirror-like objects using patterns reflected from screens surrounding them and captured with an arc of eleven cameras. Such patterned illumination is also applied to detect defects on glossy materials [22] or used for single shot surface reconstruction [20]. Similarly Tarini *et al.* [38] use multiple colored patterns, achieving high resolution and accurate reconstruction, despite using only one viewpoint. Jacquet *et al.* [15] use the deformation of straight lines reflected in the windows of buildings to recover their near planar normal maps.

Although some of these approaches achieve good reconstruction accuracy they only apply for calibrated scenes where the illumination is carefully controlled and hundreds of images are typically required. In contrast, we utilize only a handful of images and do not require any control over the environment other than it being mostly static during capture.

### 2.2. Strong diffuse components

The shape of a surface with a specular material can be recovered with controlled illumination if it has a strong diffuse component. Nehab *et al.* [24] mix a dense stereo framework with an analysis of specular consistencies under a structured illumination to recover surface normals and depth. Tunwattanapong *et al.* [39] recover the shape and normals in a controlled light environment. To deal with non-Lambertian materials and shape reconstruction Schultz [36] uses specular highlights to recover the surface viewed from different angles. That approach has been extended to specular surfaces such as [43], glossy surfaces in [8], and even materials with inter-reflections in [31]. Further, Hernández *et al.* [13] reconstruct non-Lambertian 3D objects using a multi-view photometric stereo framework. Zhou *et al.* [44] extend the idea to support spatially varying isotropic materials.

Although related in spirit to our method as they handle specularities, these methods still assume photoconsistency. Our focus is on perfectly specular surfaces, for which this assumption does not hold.

**Shape from Natural Illumination.** Johnson and Adelson [16] present a method to recover surface normals of single-color lambertian objects from a single photograph and a model of the environment in the form of a diffuse calibration sphere. Oxholm and Nishino [27] extend this method by using an environment map to fit a probabilistic model to the reflectance of the object [25]. They further expand this work to produce reconstructions from multiple calibrated images [28], which shares many similarities with our work. Most notably, they also represent the incident light field as an environment map, form probability distributions of the surface normals and iteratively refine both the surface geometry and the normals to reconstruct the object. Their method is applicable to a wide range of objects, provided that their materials either contain a strong diffuse component or rough specularities.

However, their focus is not on recovering the shape of objects exhibiting sharp specular reflections. This leaves us without a viable solution for reconstructing such objects in uncontrolled environments. In contrast, our method is tailor-made to reconstruct purely specular objects. Our model for the distribution of the surface normals is able to cope with the high frequency information observed in mirror reflections, is robust to outliers and explicitly handles inter-reflections.

### 3. Method

We aim to reconstruct a reflective object that can be characterized by its closed surface  $S$ . We assume that the object is small enough with respect to its surroundings to treat the environment as infinitely far away. We represent the incident light field as an environment map  $E(\mathbf{r})$ , *i.e.* a function that maps directions to colors, (see Figure 2). In this setting, the color of a surface point  $\mathbf{v} \in S$ , observed from the direction of the eye vector  $\mathbf{e}$ , only depends on the surface normal  $\mathbf{n}$  and the position of the camera  $I$  relative to  $\mathbf{v}$ . The

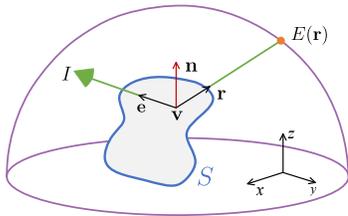


Figure 2. **Our model.** Light from the environment map  $E(\mathbf{r})$  hits vertex  $\mathbf{v}$  and is reflected about the unknown normal  $\mathbf{n}$ . The reflection follows eye vector  $\mathbf{e}$  toward camera  $I$

user obtains a set of calibrated photographs of the surface, seen from different viewpoints. We use the observed color in an image to infer the possible normals at  $\mathbf{v}$ . If every point in the environment had a unique color, a single lookup would reveal the normal explaining  $\mathbf{v}$ 's color. However, similar colors might be explained by different normals. For example, a large range of normals can explain a blue sky reflection, or observing green could mean reflecting grass or trees, introducing ambiguity. We reduce this uncertainty by combining the multiple observations of the same surface point from different viewing directions.

**Overview.** Our method proceeds as follows (see Figure 3): Our initial surface estimate  $S_e$  is a visual hull [19] formed through voxel carving from the calibrated cameras and hand made object silhouettes. A triangulated mesh of the visual hull is generated using screened Poisson surface reconstruction [18], then remeshed using [10] to ensure uniform triangle shape and size. This eliminates the need to compensate for different shapes and areas in the surface optimization. As described in Section 3.1, we build a probability distribution over the possible normals for each vertex, by comparing the observed colors against the ones in the environment. We

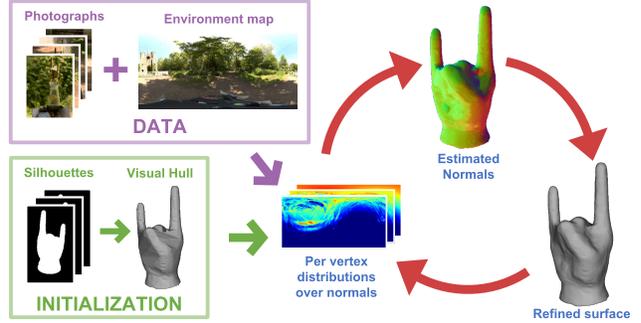


Figure 3. **Our pipeline.** The user captures multi-view images of a reflective object. The visual hull is constructed from masks of the object in each image. From this initialization and a captured environment map, our method alternates between refining the surface normals and its shape, to explain the observed reflections.

account for the estimated material's tint and inter-reflections before merging the probability distributions obtained from each of the photographs using Bayesian inference. In Section 3.2, we extract a representative normal for each vertex by assuming local smoothness, and then refine the mesh to better explain these normals. This procedure is then iterated until convergence. We also propose, in Section 3.2.2, an efficient and robust method to estimate frontier points [11, 9] on the initial mesh.

#### 3.1. Estimating normal distributions

**Model.** Given a vertex  $\mathbf{v}$  on a surface  $S_e$ , we denote  $I = \{I_k\}_{k=1 \dots K}$  as the set of cameras which  $\mathbf{v}$  is visible to and  $\mathbf{c}_{ok}$  as its observed color in the image from  $I_k$ , see Figure 2. We uniformly and densely sample the normal space  $S^2$  to form a discrete probability distribution over each vertex's normal.

We model the conditional probability of observing  $\mathbf{v}$ 's color given a normal  $\mathbf{n}$  in image  $I_k$  as a multivariate t-distribution  $t$  (Appendix A), centered on the color of the environment expected to be reflected from this viewpoint:  $\mathbf{c}_{ek}(\mathbf{n})$ . We choose the multivariate t-distribution because it is robust to outliers thanks to its heavy tails.

Given the current surface estimate, we identify two cases where reflection is impossible. First, if the normal  $\mathbf{n}$  is facing away from the observer, *i.e.* if the angle between  $\mathbf{n}$  and the eye vector  $\mathbf{e}_k$  is larger than  $90^\circ$ . Second, if the reflected vector  $\mathbf{r}_n(\mathbf{e}_k)$  crosses the local tangent plane, *i.e.* if the angle between  $\mathbf{r}_n(\mathbf{e}_k)$  and the current surface normal  $\mathbf{n}_m$  is larger than  $90^\circ$ . Interestingly, the naive solution of assigning probability zero to such normals prevents deep concavities from being recovered. To resolve this, we model the probability of observing any color as a uniform distribution  $u(\mathbf{c}_{ok})$  if the normal produces an impossible reflection. Formally,

$$P(\mathbf{c}_{ok}|\mathbf{n}) = \begin{cases} t(\mathbf{c}_{ok}|\mathbf{c}_{ek}(\mathbf{n}), \nu, \sigma^2), & \text{if } \mathbf{e}_k \cdot \mathbf{n} > 0 \\ & \text{and } \mathbf{r}_n(\mathbf{e}_k) \cdot \mathbf{n}_m > 0 \\ u(\mathbf{c}_{ok}), & \text{otherwise,} \end{cases} \quad (1)$$

where we set  $\sigma = 0.01$ ,  $\nu = 0.01$  in all our experiments. We define the support of the uniform distribution to be the volume  $[0, \hat{c}]^3$ , where  $\hat{c}$  is the largest value for all of the combined color channels of the environment map, *i.e.*  $u(\mathbf{c}_{ok}) = \hat{c}^{-3}$ .

Modeling the probability of an observed color this way assigns high likelihoods to good samples, while not completely discarding normals producing outliers. We found that outliers fit into two main categories:

- The estimated surface originating from space carving will not exhibit interior concavities. This means that a vertex  $\mathbf{v}$  far away from the true surface  $S$  is likely to give observations produced by different normals for each image, see Figure 4a. Similarly to [28], we alleviate this issue by not considering observations for which the angle between the current surface normal and the view direction is larger than  $60^\circ$ .
- When an observation is the result of an inter-reflection, see Figure 4b, the normal best explaining such a color will not correspond to the true surface normal. Later in this section we describe a method to explicitly model such observations.

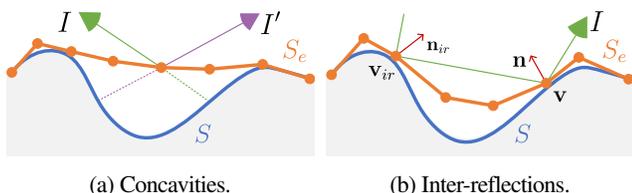


Figure 4. **Outliers.** In these cases the reflected color cannot be explained by the estimated surface or the local normal alone.

We assume each observation is independent and use Bayes’ rule to compute the posterior probability distributions over the normals at each vertex. That is,

$$P(\mathbf{n}_j|I) = P(\mathbf{n}_j) \prod_{k=1}^K \frac{P(\mathbf{c}_{ok}|\mathbf{n}_j)}{P(\mathbf{c}_{ok})}, \quad (2)$$

$$P(\mathbf{c}_{ok}) = \sum_j P(\mathbf{c}_{ok}|\mathbf{n}_j)P(\mathbf{n}_j). \quad (3)$$

We set the prior  $P(\mathbf{n}_j)$  to follow a discretized von Mises-Fisher distribution centered around the current mesh normal for each vertex  $\mathbf{v}$  and with concentration parameter  $\kappa = 25$  for both real-world and synthetic data (Appendix A). As can be seen in Figure 6, the influence of the prior decreases with the number of observations.

**Material estimation.** We model the surface material as a tinted perfect mirror with no diffuse component. In other words, we express the color  $\mathbf{c}_{ek}(\mathbf{n})$  reaching the observer as the reflected color from the environment map  $E$  scaled by an RGB triplet  $\rho$ . Formally,

$$\mathbf{c}_{ek}(\mathbf{n}) = \rho E(\mathbf{r}_n(\mathbf{e}_k)), \quad (4)$$

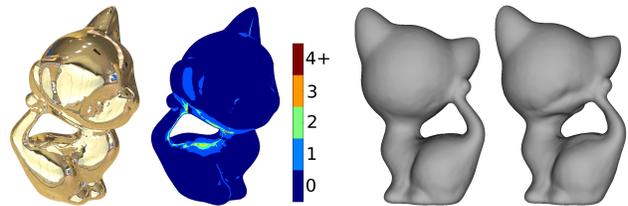


Figure 5. **Inter-reflections.** **Left:** We found that the vast majority of observations are the results of either direct reflections (dark blue) or single-bounce inter-reflections (light blue). **Right:** Ignoring inter-reflections causes artifacts (right) not present in our reconstructions (left)

where  $\mathbf{r}_n(\mathbf{e}_k)$  is the reflection of the eye vector  $\mathbf{e}_k$  about a normal  $\mathbf{n}$  at vertex  $\mathbf{v}$ . The Fresnel effect is significant only for observations at glancing angles, as we discard such observations we do not incorporate it in our material model.

To compute the material tint, we use the estimated surface to obtain the colors  $E(\mathbf{r}_n(\mathbf{e}_k))$  for each vertex in all of the photographs. We then obtain  $\rho$  by mapping these colors to the observed ones using a robust linear regression with 20 samples and  $10^4$  RANSAC iterations.

**Inter-reflections.** We classify the observations as either *direct reflections* or *inter-reflections*. A direct reflection is the result of a single mirror reflection between the environment and the observer. In contrast, an inter-reflection is the result of two or more perfect mirror reflections on the surface of the object, see Figure 4b. These observations are difficult to model because instead of depending on a single surface normal, an arbitrary number of surface normals may be needed to explain them.

However, we observe that, as seen in Figure 5, most of the observations can be modeled as either a direct reflection (*dr*) or a single bounce inter-reflection (*ir*). Based on this observation, we focus on the first level of inter-reflection only. A simple but naive approach is to ignore an observation if it would be an inter-reflection given the current estimate of the surface. This would however be a waste of information. We instead choose to augment  $\mathbf{c}_{ek}$  to explicitly take inter-reflections into account by ray-tracing the estimated surface  $S_e$  and computing the colors resulting from one bounce reflections. In other words,

$$\mathbf{c}_{ek}(\mathbf{n}) = \begin{cases} \rho E(\mathbf{r}_n(\mathbf{e}_k)) & (dr) \\ \rho^2 E(\mathbf{r}_{\mathbf{n}_{ir}(k)}(\mathbf{r}_n(\mathbf{e}_k))) & (ir), \end{cases} \quad (5)$$

where  $\mathbf{n}_{ir}(k)$  is the normal where the reflected ray intersects with the mesh again (see Figure 4b).

Even though this method is an approximation because of the use of the estimated surface  $S_e$ , its accuracy increases with the number of iterations as the surface gets refined. Figure 5 shows that it significantly improves the results and that an artifact-free reconstruction would not be possible without explicitly handling inter-reflections.

## 3.2. Optimization

We use the estimated probability distributions for the surface normals to drive the shape of the object. Taking heed of the experimental evaluation performed by Hernandez *et al.* [41], we do not directly refine the mesh based on the probability distributions, but adopt a two-step process instead.

We first extract representative normals  $\mathbf{n}_i$  for each vertex  $\mathbf{v}_i$  on the surface while enforcing local smoothness (Section 3.2.1). We then refine the surface to better explain these normals (Section 3.2.2). To adequately constrain this problem, we make use of frontier points (Section 3.2.3) whose positions are guaranteed to match those of the original object. We iterate this process until convergence, making use the refined surface at each iteration to produce more confident probability distributions for the surface normals.

### 3.2.1 Representative normals

We fix the location of all vertices and model the extraction of a representative normal  $\mathbf{n}_i$  for every vertex  $\mathbf{v}_i$  as an energy minimization problem. Its objective function consists of two distinct parts,

$$E_n = \alpha_d E_d + \alpha_s E_s, \quad (6)$$

where  $E_d$  is the data term and  $E_s$  is the smoothness term. We set  $\alpha_d = 1$ ,  $\alpha_s = 5 \times 10^5$  and initialize each normal to the global maximum of its corresponding distribution in all our experiments.

**Data term.** The first term penalizes normals that have low probability according to the estimated distributions. Specifically,

$$E_d = \sum_i |\log P(\mathbf{n}_i)|^2. \quad (7)$$

As the distributions of the normals are discretized, we remodel the continuous density in log-space by placing von Mises-Fisher distributions at each discrete probability sample  $P(n_p)$ . Namely,

$$\log P(\mathbf{n}_i) \propto \sum_p \log P(n_p) e^{\kappa \mathbf{n}_p^T \mathbf{n}_i}, \quad (8)$$

where we set  $\kappa = \theta_d^{-2}$ . For large values of  $\kappa$ , the angle  $\theta_d$  roughly corresponds to one standard deviation (Appendix A). We initialize  $\theta_d$  to  $5^\circ$  and linearly decrease it to reach  $2^\circ$ , its minimum value, at the fifth iteration of the entire optimization process.

**Smoothness term.** The second term ensures local smoothness of the normals, while not penalizing curvature. This formulation avoids the caveats of a simple per-edge dot product, which would only yield zero cost for flat surfaces. Instead our

cost depends on the mean of the normal vectors in the one-ring neighborhood. In other words,

$$E_s = \sum_i \left| 1 - \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \mathbf{n}_i \cdot \mathbf{n}_j \right|^2, \quad (9)$$

where  $\mathcal{N}_i$  is the one-ring neighborhood of  $\mathbf{v}_i$ .

### 3.2.2 Surface refinement

We refine the surface to better explain the now fixed representative normals. Again, we model this problem as the minimization of an objective function consisting of three distinct parts. Namely,

$$E_m = \alpha_{\text{mesh}} E_{\text{mesh}} + \alpha_v E_v + \alpha_{\text{fp}} E_{\text{fp}}, \quad (10)$$

where  $E_{\text{mesh}}$  is the normal term,  $E_v$  is the volume term and  $E_{\text{fp}}$  is the frontier point term. We set  $\alpha_{\text{mesh}} = 10$ ,  $\alpha_v = 1$  and  $\alpha_{\text{fp}} = 1$  in all our experiments. For these parameters to work across multiple scales we resize the mesh for the surface refinement so that the average edge size is unit length. Furthermore, inspired by [23] we only allow for vertices to be displaced along their representative normals during the optimization to minimize the occurrence of self-intersecting triangles.

**Normal term.** For the purpose of matching the mesh to the optimized normals we employ the linear cost  $E_{\text{mesh}}$  from [23]. This cost enforces the edges in the one-ring neighborhood  $\mathcal{N}_i$  around each vertex  $\mathbf{v}_i$  to be perpendicular to its estimated normal  $\mathbf{n}_i$ . Formally,

$$E_{\text{mesh}} = \sum_i \sum_{j,k \in \mathcal{N}_i} |\mathbf{n}_i \cdot (\mathbf{v}_j - \mathbf{v}_k)|^2. \quad (11)$$

**Volume term.** We introduce a penalty to prevent the mesh from increasing in volume compared to the visual hull. We define this penalty as

$$E_v = \sum_i |\max(0, \mathbf{n}_i \cdot (\mathbf{v}_i - \bar{\mathbf{v}}_i))|^2, \quad (12)$$

where  $\bar{\mathbf{v}}_i$  is the vertex closest to  $\mathbf{v}_i$  on the visual hull.

**Frontier point term.** In epipolar geometry *contour generators* [11, 9] are the set of 3D curves, one per camera, where the ground truth surface coincides with the visual hull. Each of these curves projects onto the silhouette contour in its associated image. Points which lie at the intersection of two contour generators are *frontier points*. If found, such points can be used as unbiased and strong constraints in the surface optimization to anchor the location of vertices and avoid drift. Based on this, we penalize vertices close to frontier points for deviating from their initial visual hull positions. In other words,

$$E_{\text{fp}} = \sum_i \|\mathbf{v}_i - \bar{\mathbf{v}}_i\|^2 e^{-r_i^2 / 2\sigma_{\text{fp}}^2}, \quad (13)$$

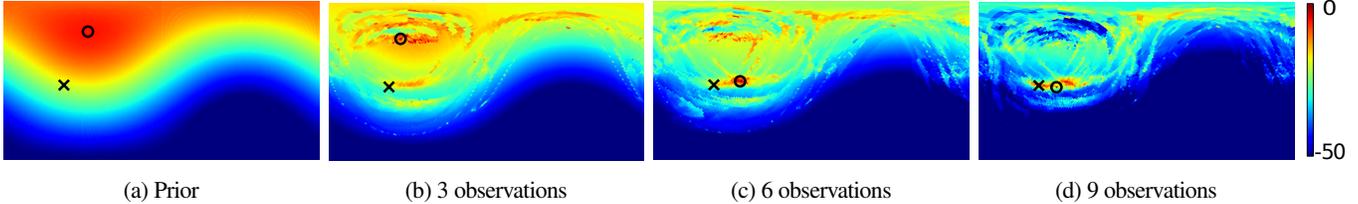


Figure 6. The evolution of the probability distribution (seen in log-space) over the possible normals for a vertex. The prior is a von Mises-Fisher distribution centered on the current surface normal. As the vertex is observed in more images, we see that the distribution becomes more confident and its mode (the circle) approaches the ground truth normal (the cross).

where  $r_i$  is the distance from  $\mathbf{v}_i$  to the closest point in the set of frontier points and  $\sigma_{fp} = 3$ .

### 3.2.3 Estimating frontier points

As our initial surface estimate is in the form of a visual hull, we cannot rely on its contours alone to identify frontier points. Indeed, points on the visual hull that project onto the contour of a given image  $I$  form a superset of the associated contour generator. For smooth surfaces, a frontier point  $\mathbf{v}$  associated with two cameras  $I$  and  $I'$  has its normal orthogonal to the epipolar plane defined by the triplet  $(\mathbf{v}, I, I')$ . A naive method would identify a vertex as a frontier point if it lies on the contours in two images and if its normal is orthogonal to the corresponding epipolar plane. However, we found this method unreliable and prone to false positives due to approximate contours and inaccurate initial normals.

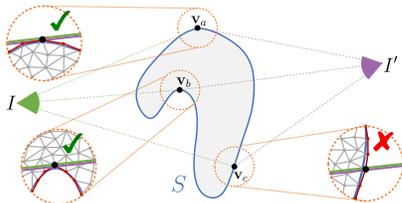


Figure 7. **Frontier points.** The vertices  $\mathbf{v}_a, \mathbf{v}_b$  and  $\mathbf{v}_c$  all lie on the contours in the images taken from  $I$  and  $I'$ . Only  $\mathbf{v}_c$  is not identified as a frontier point as the epipolar plane (the purple-green line) crosses the union of the projective neighborhoods  $\mathcal{N}_c(I)$  and  $\mathcal{N}_c(I')$  (highlighted in red).

Instead, we observe that a plane intersecting the surface at a vertex  $\mathbf{v}$ , but no other point in its neighborhood, is a tangent plane at  $\mathbf{v}$ . Based on this observation, we identify a vertex  $\mathbf{v}_i$  that lies on the contour in two images  $I$  and  $I'$  as a frontier point if its neighborhood does not cross the epipolar plane (Figure 7). To cope with inaccurate contours, we let the neighborhood of  $\mathbf{v}_i$  depend on  $I$  and  $I'$ . Specifically, we define the neighborhood of  $\mathbf{v}_i$  as the union of the *projective neighborhoods*  $\mathcal{N}_i(I)$  and  $\mathcal{N}_i(I')$ , where  $\mathcal{N}_i(I)$  is the set of points that lie on the contour from  $I$  and are inside a narrow cone with radius  $r_{fp}$  at  $\mathbf{v}_i$ . We set  $r_{fp}$  to be 7.5 times the average edge length in all our experiments.

### Post-processing

Once the surface has been refined, the assumption of a uniform triangulation no longer holds. Indeed, our energy terms do not prevent irregular triangle shapes and inverted faces. Modifying the terms to prevent such issues introduces more parameters and did not converge to the right solution in our experiments. Instead, we correct the issues in a post-processing step identical to how we obtain the triangulated mesh from the visual hull.

## 4. Implementation and results

**Data capture** We generated our synthetic dataset by rendering 36 views of four different objects, KITTY, BUNNY, PLANCK and ROCKERARM. Each object was rendered using three different high-dynamic-range (HDR) environment maps TOKYO, PAPERMILL and LOFT, all freely available online [6]. To investigate the effect of breaking the assumption that the environment is infinitely far away, we also rendered the objects in the 3D scene SPONZA using a global illumination path tracer.

Our real-world dataset contains three objects, PIGGY, TEDDY and HEAVYMETAL captured in two scenes, THEATER and CHURCH. Capturing all objects in one scene took approximately 50 minutes on site. We obtained the images for this dataset by merging bracketed DSLR photographs into linear HDR images. The environment map was obtained by stitching together wide-angle photographs. We manually created the silhouette images using standard image editing software, which took roughly one hour per object. We also acquired accurate geometry for PIGGY and TEDDY by coating them with diffusing spray and laser scanning them. The camera extrinsics and intrinsics were recovered using a standard structure from motion package [21] and we attached texture rich postcards to the tripod supporting the objects to ensure good calibration. We aligned the environment map and the structure from motion coordinate space in less than ten minutes by manually locating four point correspondences in the environment map and the background of the object’s photographs and then applying Kabsch’s algorithm [17] to align them.

**Optimization.** We use the HEALpix projection [12] to uniformly sample the normal space  $S^2$  with  $N_p = 12 \times 4^{N_r}$  samples. Unless mentioned otherwise we use the resolution exponent  $N_r = 5$  and  $N_p = 12288$  samples. We downsample the en-

vironment map to match the sampling density of  $S^2$ . The distributions are computed on a GPU and the Optix Prime ray tracing library [29] is used to resolve visibilities and inter-reflections.

We use the Ceres Solver library [4] implementation of the L-BFGS algorithm to both extract representative normals and refine the surface estimate. The optimizations are run with an unbounded number of iterations until the convergence criterion  $d\mathbf{E}/\mathbf{E} < 10^{-6}$  is met. In the pre- and post-processing stages we remesh all our surfaces to contain 50000 vertices.

### 4.1. Synthetic results

To test our method and validate the assumptions it relies on, we ran a set of experiments using our synthetic dataset. See Figure 10 for qualitative results of our method on a cross section of the synthetic dataset.

**Environments and inter-reflections.** In Figure 9 we see that the choice of environment only weakly affects the reconstruction quality. In the same figure, we also analyze the impact of inter-reflections on the reconstruction quality. As can be seen from the convergence plots, ignoring inter-reflections often causes the reconstruction to diverge. Discarding observations that would be inter-reflections based the current surface estimate wastes useful data and does not reach the reconstruction quality of our explicit method.

**Proximity.** To evaluate the impact of the object’s proximity to the environment on the reconstruction quality, we rendered two new versions of BUNNY in SPONZA. One where it is twice as large and one where it is twice as small. As the size of the scene remains constant, the visual angle of the object in the images increases with the size of the object as the observer is unable to move further away. Intuitively, the reconstruction of the larger object should be of lower quality as the reflections disagree more with the environment map due to parallax. In our other experiments, the visual angle of BUNNY is  $10^\circ$ . As expected, the reconstruction error of the small version ( $5^\circ$ ) is 7.4% lower while the error increases by 8.3% for the large version ( $20^\circ$ ).

**Resolution.** Table 1 shows that the final reconstruction error is not heavily affected by the choice of the sampling density of

the normal space  $S^2$ . In our other experiments, we use  $N_r = 5$  as it is a good compromise between speed and quality.

Resolution exponent $N_r$					
2	3	4	5	6	7
-21.74%	-31.22%	-31.24%	-31.86%	-32.58%	<b>-32.8%</b>

Table 1. The decrease in RMS error compared to the visual hull for BUNNY in SPONZA when varying  $N_r$  for the sampling density of  $S^2$ .

### 4.2. Real-world results

We also evaluate our method on a real-world dataset, see Figure 11 for qualitative results. Despite the challenges of real-world capture, our method accurately reconstructs the objects. Our method recovers deep concavities such as the palm of HEAVYMETAL and fine details such as the eyes of PIGGY. Results from the CHURCH scene show that it is hard to recover from false positive frontier points. Scene specific parameters could alleviate this, but we used the same parameters in all experiments. See Figure 1 for our reconstruction of TEDDY in the INCEPTION scene and the video in the supplemental material for more real-world results.

**Comparison with [28].** Although Oxholm and Nishino [28] do not explicitly target highly specular objects, their BRDF model does in theory support this case. Unfortunately, we cannot directly compare our methods as they were unable to run their method on new datasets and it is not possible to run our method on their dataset due to poorly calibrated images.

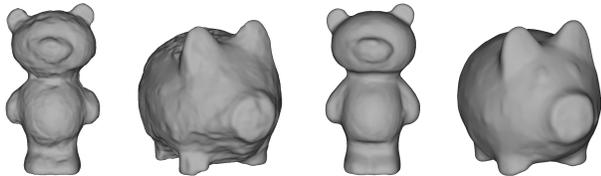
Instead, we modified our method to compute the distributions over the surface normals using their algorithm, keeping the rest of our pipeline identical. This compromise was made as using their area priors during surface refinement led to worse reconstructions (Appendix B). Figure 8 shows that this introduces bumps on the reconstructed surfaces, because their method is sensitive to outliers without a BRDF that smooths out the high frequency components in the environment.

## 5. Conclusion

We have presented a method to accurately recover the shape of highly specular objects in uncontrolled environments from a small number of photographs. To our knowledge, our method is the first to achieve this using only commodity hardware.

Our method has some limitations. It requires handmade silhouettes of the object in the input photographs, which introduces the risk of false positive frontier points as these are deterministically constructed from the silhouettes and our optimization framework does not always recover from such points. Interesting avenues for future work would be to partially or fully automate the silhouette extraction and improve the silhouette constraints in the surface optimization.

Finally, to encourage comparison and further work, source code and datasets are available on the project webpage.



(a) Our interpretation of [28]

(b) Our method

Figure 8. Reconstructions of TEDDY and PIGGY in THEATER after 5 iterations of either our original method or a modified version that uses [28] to compute the distributions over the surface normals.

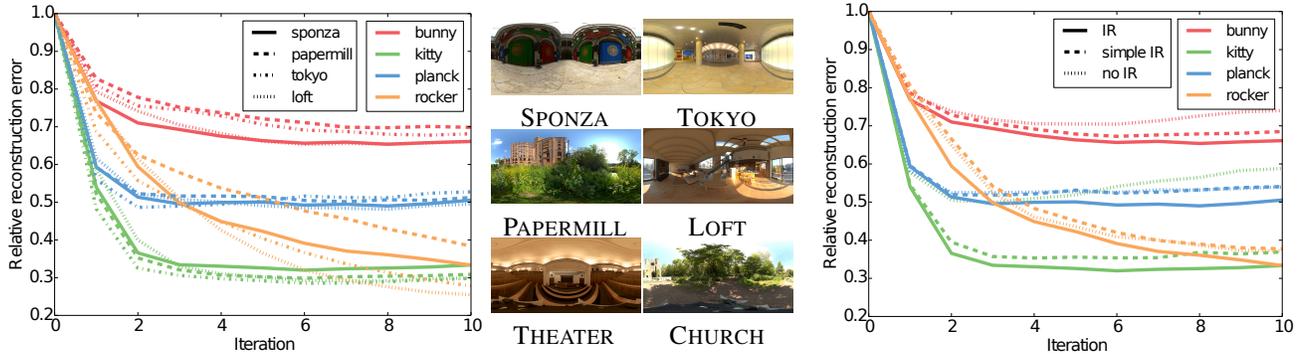


Figure 9. **Left.** The RMS error relative to that of the initial visual hull for four synthetic objects, each rendered in four environments. Our method is robust to varied environments and converges in under 10 iterations. **Middle.** The panoramic image of the four synthetic environments we used in our experiments, as well as the two environments in which we captured our real-world objects. **Right.** The RMS error relative to that of the visual hull for different strategies that deal with inter-reflections: Ignoring inter-reflections all together (no IR), discarding observations that inter-reflect given the estimated surface (simple IR) and our explicit method (IR).

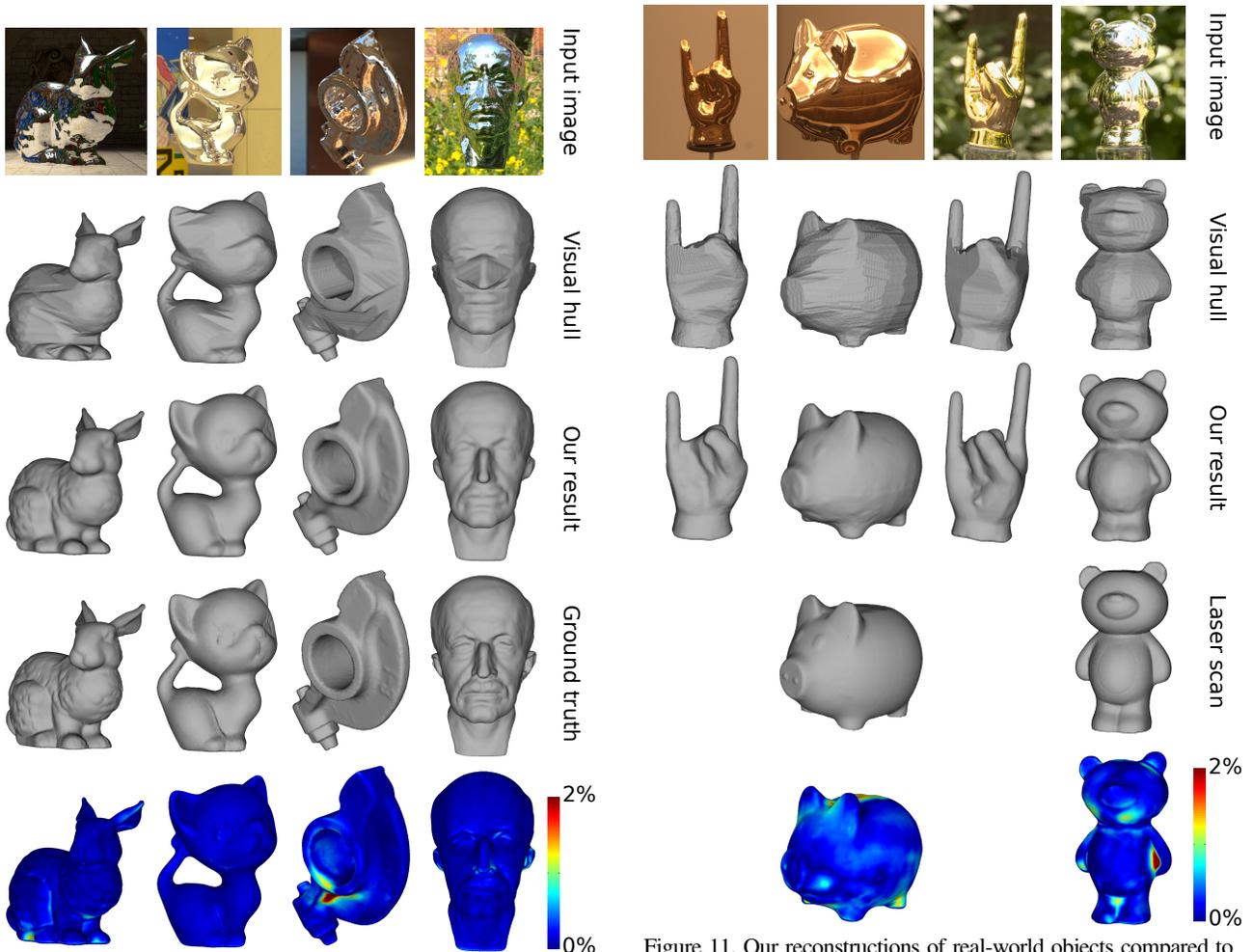


Figure 10. Our reconstructions of synthetic objects compared to one of the input photographs and the ground truth mesh. **Left to right:** BUNNY in the SPONZA scene, KITTY in TOKYO, ROCKERARM in LOFT and PLANCK in PAPERMILL. All objects were reconstructed from 36 input photographs. The last row shows the reconstruction errors relative to the bounding box diagonals.

Figure 11. Our reconstructions of real-world objects compared to an input photograph and a laser scan (where applicable). **Left to right:** HEAVYMETAL and PIGGY in the THEATER scene (35 and 25 input photographs respectively), HEAVYMETAL and TEDDY in the CHURCH scene (27 and 26 input photographs respectively). The last row shows the reconstruction errors relative to the bounding box diagonals.

**Acknowledgements.** We thank Moos Huetting, Tara Ganepola, Corneliu Iliescu and Aron Monszpart for their invaluable help to get the first version of this paper ready. We also thank George Drettakis and the UCL Vision Group for their helpful comments. The authors are supported by the UK EPSRC-funded Eng. Doctorate Centre in Virtual Environments, Imaging and Visualisation (EP/G037159/1), EU project CR-PLAY (no 611089) [www.cr-play.eu](http://www.cr-play.eu), and EPSRC EP/K023578/1.

## References

- [1] Y. Adato and O. Ben-Shahar. Specular flow and shape in one shot. In *BMVC*, 2011.
- [2] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Toward a theory of shape from specular flow. In *ICCV*, 2007.
- [3] Y. Adato, T. Zickler, and O. Ben-Shahar. Toward robust estimation of specular flow. In *BMVC*, 2010.
- [4] S. Agarwal, K. Mierle, and Others. Ceres solver. <http://ceres-solver.org>.
- [5] J. Balzer, S. Hofer, and J. Beyerer. Multiview specular stereo reconstruction of large mirror surfaces. In *CVPR*, 2011.
- [6] C. Bloch. sIBL archive. [www.hdrlabs.com/sibl/archive.html](http://www.hdrlabs.com/sibl/archive.html).
- [7] T. Bonfort and P. Sturm. Voxel carving for specular surfaces. In *ICCV*, 2003.
- [8] H.-S. Chung and J. Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. In *CVPR*, 2008.
- [9] R. Cipolla and P. Giblin. *Visual motion of curves and surfaces*. Cambridge University Press, 2000.
- [10] S. Fuhrmann, J. Ackeremann, T. Kalbe, and M. Goesele. Direct resampling for isotropic surface remeshing. In *VMV*, 2010.
- [11] P. Giblin, F. E. Pollock, and J. Rycroft. Recovery of an unknown axis of rotation from the profiles of a rotating surface. *JOSA A*, 1994.
- [12] K. M. Gorski, E. Hivon, A. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2), 2005.
- [13] C. Hernández, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *PAMI*, 2008.
- [14] I. Ihrke, K. N. Kutulakos, H. Lensch, M. Magnor, and W. Heidrich. Transparent and specular object reconstruction. *CGF*, 29(8), 2010.
- [15] B. Jacquet, C. Häne, K. Köser, and M. Pollefeys. Real-world normal map capture for nearly flat reflective surfaces. In *ICCV*, 2013.
- [16] M. K. Johnson and E. H. Adelson. Shape estimation in natural illumination. In *CVPR*, 2011.
- [17] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5), 1976.
- [18] M. Kazhdan and H. Hoppe. Screened Poisson surface reconstruction. *ACM TOG*, 32(1), 2013.
- [19] A. Laurentini. The visual hull concept for silhouette-based image understanding. *PAMI*, 16(2), 1994.
- [20] M. Liu, R. Hartley, and M. Salzmann. Mirror surface reconstruction from a single image. In *CVPR*, 2013.
- [21] P. Moulon, P. Monasse, and R. Marlet. Adaptive structure from motion with a contrario model estimation. In *ACCV*. 2012.
- [22] T. Nagato, T. Fuse, and T. Koezuka. Defect inspection technology for a gloss-coated surface using patterned illumination. In *IS&T/SPIE Electronic Imaging*, 2013.
- [23] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM TOG*, 24(3), 2005.
- [24] D. Nehab, T. Weyrich, and S. Rusinkiewicz. Dense 3D reconstruction from specular consistency. In *CVPR*, 2008.
- [25] K. Nishino. Directional statistics BRDF model. In *ICCV*, 2009.
- [26] M. Oren and S. K. Nayar. A theory of specular surface geometry. *IJCV*, 24(2), 1997.
- [27] G. Oxholm and K. Nishino. Shape and reflectance from natural illumination. In *ECCV*, pages 528–541, 2012.
- [28] G. Oxholm and K. Nishino. Multiview shape and reflectance from natural illumination. In *CVPR*, pages 2163–2170, June 2014.
- [29] S. G. Parker, J. Bigler, A. Dietrich, H. Friedrich, J. Hoberock, D. Luebke, D. McAllister, M. McGuire, K. Morley, A. Robison, et al. Optix: a general purpose ray tracing engine. *ACM TOG*, 29(4):66, 2010.
- [30] S. Roth and M. J. Black. Specular flow and the recovery of surface structure. In *CVPR*, 2006.
- [31] R. Ruiters and R. Klein. Heightfield and spatially varying BRDF reconstruction for materials with interreflections. *CGF*, 28(2), 2009.
- [32] A. C. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. Agrawal. Image invariants for smooth reflective surfaces. In *ECCV*, 2010.
- [33] S. Savarese, M. Chen, and P. Perona. Recovering local shape of a mirror surface from reflection of a regular grid. In *ECCV*, 2004.
- [34] S. Savarese, M. Chen, and P. Perona. Local shape from mirror reflections. *IJCV*, 64(1), 2005.
- [35] S. Savarese, L. Fei-Fei, and P. Perona. What do reflections tell us about the shape of a mirror? *ACM TOG*, 2004.
- [36] H. Schultz. Retrieving shape information from multiple images of a specular surface. In *PAMI*, volume 16, 1994.
- [37] M. F. Tappen. Recovering shape from a single image of a mirrored surface from curvature constraints. In *CVPR*, 2011.
- [38] M. Tarini, H. P. Lensch, M. Goesele, and H.-P. Seidel. *3D acquisition of mirroring objects*. MPI Informatik, Bibliothek & Dokumentation, 2003.
- [39] B. Tunwattanapong, G. Fyffe, P. Graham, J. Busch, X. Yu, A. Ghosh, and P. Debevec. Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM TOG*, 32(4), 2013.
- [40] Y. Vasilyev, T. Zickler, S. Gortler, and O. Ben-Shahar. Shape from specular flow: Is one flow enough? In *CVPR*, 2011.
- [41] G. Vogiatzis, C. Hernandez, and R. Cipolla. Reconstruction in the round using photometric normals and silhouettes. In *CVPR*, 2006.
- [42] M. Weinmann, A. Osep, R. Ruiters, and R. Klein. Multi-view normal field integration for 3D reconstruction of mirroring objects. In *ICCV*, 2013.
- [43] Z. Zheng, M. Lizhuang, L. Zhong, and Z. Chen. An extended photometric stereo algorithm for recovering specular object shape and its reflectance properties. *ComSIS*, 7(1), 2010.
- [44] Z. Zhou, Z. Wu, and P. Tan. Multi-view photometric stereo with spatially varying isotropic materials. In *CVPR*, 2013.