Scene reassembly after multimodal digitization and pipeline evaluation using photorealistic rendering

Jonathan Dyssel Stets,[1,†] Alessandro Dal Corso,[1,†] Jannik Boll Nielsen,[1] Rasmus Ahrenkiel Lyngby,[1] Sebastian Hoppe Nesgaard Jensen,[1] Jakob Wilm,[1] Mads Brix Doest,[1] Carsten Gundlach,[2] Eythor Runar Eiriksson,[1] Knut Conradsen,[1] Anders Bjorholm Dahl,[1] Jakob Andreas Bærentzen,[1] Jeppe Revall Frisvad,[1,*] and Henrik Aanæs[1]

[1]*Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, 2800 Kongens Lyngby, Denmark*
[2]*Department of Physics, Technical University of Denmark, Fysikvej, 2800 Kongens Lyngby, Denmark*
*Corresponding author: jerf@dtu.dk

Transparent objects require acquisition modalities that are very different from the ones used for objects with more diffuse reflectance properties. Digitizing a scene where objects must be acquired with different modalities requires scene reassembly after reconstruction of the object surfaces. This reassembly of a scene that was picked apart for scanning seems unexplored. We contribute with a multimodal digitization pipeline for scenes that require this step of reassembly. Our pipeline includes measurement of bidirectional reflectance distribution functions and high dynamic range imaging of the lighting environment. This enables pixelwise comparison of photographs of the real scene with renderings of the digital version of the scene. Such quantitative evaluation is useful for verifying acquired material appearance and reconstructed surface geometry, which is an important aspect of digital content creation. It is also useful for identifying and improving issues in the different steps of the pipeline. In this work, we use it to improve reconstruction, apply analysis by synthesis to estimate optical properties, and to develop our method for scene reassembly.   © 2017 Optical Society of America

1. INTRODUCTION

Several research communities work on techniques for optical acquisition of physical objects and their appearance parameters [1–5]. Thus, we are now able to acquire nearly any type of object and perform a computer graphics rendering of nearly any type of scene. The range of applications is broad and includes movie production [2], cultural heritage preservation [3], 3D printing [4], and industrial inspection [5]. A gap left by these multiple endeavors is a coherent scheme for acquiring a scene consisting of several objects that have very different appearance parameters, together with the reassembly of a digital replica of such a scene. Our objective is to fill this gap for the combination of transparent and opaque objects, as many real-world scenarios exhibit this combination. An example is a living room, like the one rendered in Fig. 1 (right). We propose a pipeline for acquiring and reassembling digital scenes from this type of heterogeneous real-world scene. In addition, our pipeline closes

the loop by rendering calibrated images of the digital scene that are commensurable with photographs of the original physical scene (see Fig. 1, left). This allows for validation and fine-tuning of appearance parameters. The quantitative evaluation we get from pixelwise comparison of rendered images with photographs is a great improvement with respect to validation of the acquired digital representation of the physical objects.

When addressing the problem of acquiring a heterogeneous scene, there is an infinite variety of scenes and object types to choose from. So, to make our task feasible, we focus on scenes that combine glassware and non-transparent materials, more specifically, a white tablecloth and cardboard with a checkerboard pattern. We made these choices, as glass requires a different acquisition modality, the tablecloth bidirectional reflectance distribution function (BRDF) is spatially uniform but not necessarily simple, and the cardboard has simple two-color variation. The latter is particularly useful for observing how light

Fig. 1. To the left, we compare rendered images (top) with photographs (bottom). More views are available in Appendix A. The scenes to the left were digitized using our pipeline and include both glass objects and non-transparent objects (tablecloth and backdrop). To the right, we exemplify the use of our pipeline for virtual product placement using our digitized glass objects, with estimated optical properties and artifact-reduced removal of markers.

refracts through the glass. The chosen case is also of particular interest, since glass is present in many intended applications of optical 3D acquisition. Considering the highly multidisciplinary nature of our work, we have released our dataset (http://eco3d.compute.dtu.dk/pages/transparency). This facilitates further investigation by other researchers of the different steps of our pipeline with the possibility of a quantitative feedback at the end of the process.

A. Related Work and Contributions

Researchers occasionally compare renderings with photographs to provide a qualitative verification of a presented rendering technique. The work by Phong [6], Goral *et al.* [7], and Takagi *et al.* [8] are early examples of this trend. A procedure to bring a rendered image close to a photograph was first presented by Meyer *et al.* [9]. In this work, likeness of images was evaluated perceptually by human observers. Pixelwise comparison of photographs with rendered images is surprisingly uncommon. The few examples we have found are by Rushmeier *et al.* [10], Karner and Prantl [11], Pattanaik *et al.* [12], and Jones and Reinhart [13,14]. These examples build on the rendering framework described by Greenberg *et al.* [15]. Employing such a framework for more complex scenes is a long and tedious process [16]. The key issue is that a scene specification is expected as an input.

Several problems arise as a result of not having correspondence between the physical and the digital scene. Misalignment due to inaccurate scene and viewing geometry and inaccurate orientation of the lighting environment are some of the essential problems identified in previous work [17,18]. One way to deal with this problem is to calculate error for image patches when evaluating results [13,19,20]. As opposed to this, our digitization pipeline (Fig. 2) provides both reference photographs and correspondingly calibrated scene and viewing geometry so that pixelwise comparison becomes meaningful.

Pixelwise comparison of rendered images with photographs is not only useful for quantifying the photorealism of a rendering in terms of error measurements. We find it particularly useful for improving the digitization pipeline. The fact that our pipeline enables quantitative evaluation led us to more specific contributions in its different steps. These contributions are mostly in the reassembly and are as follows. (a) A cross-modality marker-based placement approach, enabling accurate placement of objects scanned with one modality into scenes scanned with another modality. (b) A soft object deformation technique dealing with surface intersections after object placement, which is critical for scenes containing transparent or translucent objects. (c) A micropolygon labeling approach for assigning BRDFs to acquired geometry. (d) A color calibration scheme enabling use
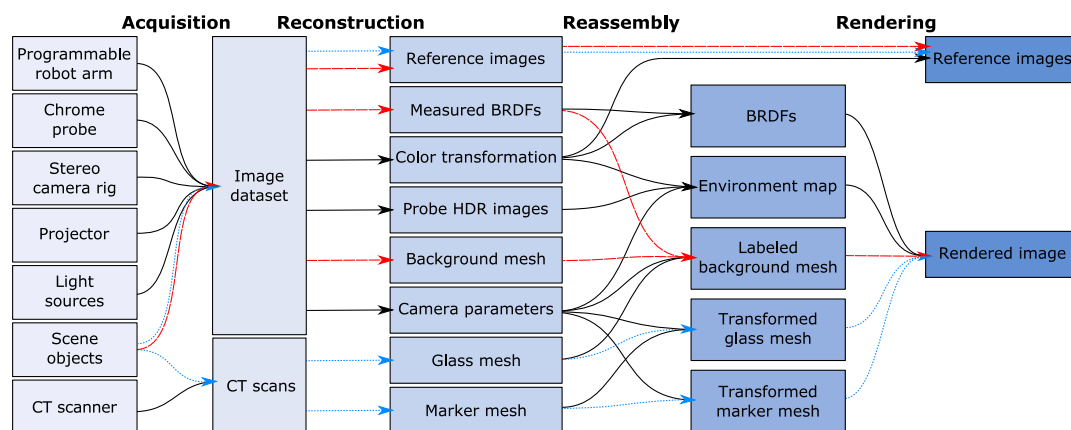

Fig. 2. Overview of our digitization pipeline in four main stages: acquisition, reconstruction, reassembly, and rendering. A video presentation of our pipeline is available in supplementary Visualization 1. Colored arrows show the path through the pipeline of transparent objects (dotted blue) and non-transparent objects (dashed red).

of spectral optical properties for calculating reflectance, transmittance, and absorption. (e) Perspective unwrapping of mirror probe images to improve precision when the environment is not very distant. (f) Use of analysis by synthesis for fine-tuning physics-based optical properties.

Digitization is most often unimodal and tailored toward objects with a specific type of surface reflectance behavior [1]. While unimodal techniques are becoming more versatile [21–23], objects with a transparent material, such as glass, still pose challenging problems. Their reflectance behavior is so different that they require an entirely different modality, such as computed tomography (CT) [24]. The transparent object must then be removed from the scene to be scanned elsewhere. In the meantime, the surrounding scene can be scanned with a more common technique. However, as the transparent object takes most of its appearance from its surroundings, it must be repositioned in the surrounding scene (physically and digitally) if we are to take reference images for comparison with rendered images. The purpose of our scene reassembly is to address this type of issue.

Our digitization technique is multimodal. Currently, such techniques seem to exist only in the context of sensor fusion [25–27]. Here, the goal is to optimize reconstruction by fusing data from different sensor modalities with complementary characteristics. Even so, the different modalities see the same object and thus work for materials with a similar reflectance behavior. The challenge is then mostly in registration of the scans. In their final remarks and suggestions for future work, Weinmann and Klein [1] discuss possible ways of combining multiple techniques tailored to different types of surface reflectance. Our pipeline is a different way to take a step in this direction.

In summary, our work makes it possible to perform multimodal digitization and scene reassembly in such a way that rendered images of the reassembled scene can be quantitatively compared to photographs of the original. This enables us to provide the first empirically founded investigation of the appearance accuracy of objects digitized using a non-optical scanner.

2. DIGITIZATION PIPELINE

We divide our pipeline into four stages: (1) acquisition, (2) reconstruction, (3) reassembly, and (4) rendering. Figure 2

provides an overview. As illustrated, transparent objects (dotted blue arrows) and non-transparent objects (dashed red arrows) take different paths through the pipeline. The acquisition stage includes structured light scanning of non-transparent objects, CT scanning of transparent objects, gonioreflectometric reflectance measurements, and photographic capture of environment, color chart, and scene reference images. Figure 3 provides details of our workflow in these acquisition steps (except the simpler captures of environment and color chart). The second stage includes reconstruction of surface meshes, material BRDFs, and color space. The third stage is reassembly of the digital scene consisting of geometric objects, material appearance properties, and environment map. The fourth and final stage is rendering and comparison with reference images.

Our acquisition stage requires an elaborate hardware setup. We assemble the physical scene in a black light-proof enclosure. This has five LED light tubes for scene lighting, which we capture by high dynamic range (HDR) imaging of a light probe. To acquire non-transparent geometry inside this enclosure, we use a structured light scanner consisting of a toe-in stereo camera rig and a light projector mounted on a robotic arm [28,29]. We chose a converging camera configuration (toe-in) to increase the overlap of the fields of view so that we get a denser point cloud per stereo view. Together with an LED-based illumination arc, we also use this camera rig with exact control for measuring isotropic BRDFs. For transparent objects, we use a CT scanner. In the following subsections, we describe the individual steps of the pipeline with a focus on details required for reproducibility and on non-standard techniques that we introduce.

A. Camera Calibration and Settings

The camera system is calibrated using a standard technique [30]. Our calibration board is an 11 by 12 black-and-white checkerboard. For the intrinsic calibration (Pass 1 of Fig. 3, left), we include a large variety of views to estimate good lens distortion coefficients. To facilitate stereo calibration, we also ensure that both cameras have the calibration board fully in view. For extrinsic calibration (Pass 2 of Fig. 3, left), we balance good coverage of the scene and good coverage of the calibration board. Since we cannot change the camera system while collecting data, we chose a small aperture to ensure that background
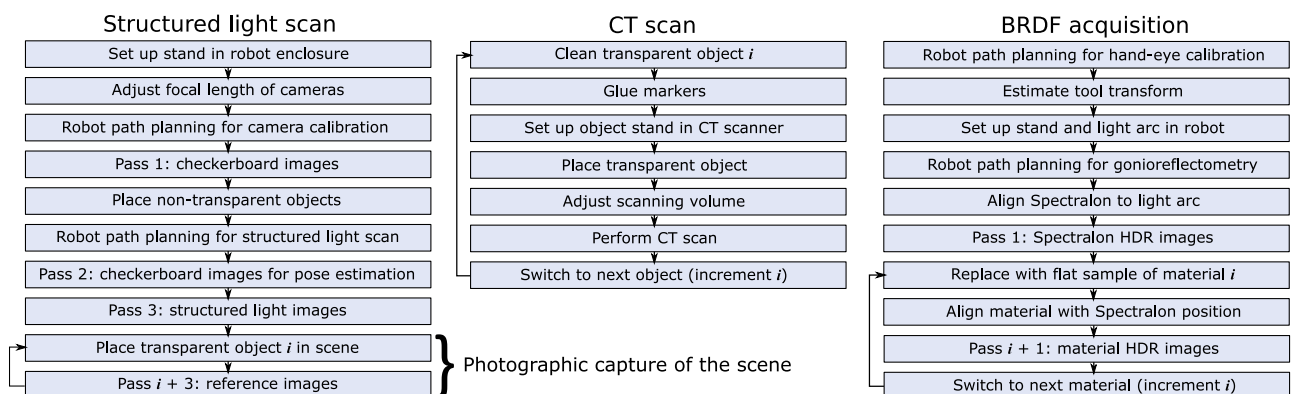

Structured light scan	CT scan	BRDF acquisition
Set up stand in robot enclosure	Clean transparent object *i*	Robot path planning for hand-eye calibration
Adjust focal length of cameras	Glue markers	Estimate tool transform
Robot path planning for camera calibration	Set up object stand in CT scanner	Set up stand and light arc in robot
Pass 1: checkerboard images	Place transparent object	Robot path planning for gonioreflectometry
Place non-transparent objects	Adjust scanning volume	Align Spectralon to light arc
Robot path planning for structured light scan	Perform CT scan	Pass 1: Spectralon HDR images
Pass 2: checkerboard images for pose estimation	Switch to next object (increment *i*)	Replace with flat sample of material *i*
Pass 3: structured light images		Align material with Spectralon position
Place transparent object *i* in scene		Pass *i* + 1: material HDR images
Pass *i* + 3: reference images		Switch to next material (increment *i*)

} Photographic capture of the scene

Fig. 3. Our workflow for scanning the geometry of non-transparent objects and collecting reference images (left), for scanning the geometry of transparent objects (middle), and for measuring material reflectance properties (right).

and projected structured light patterns are always in focus from all views. The full setup is in a dark room environment to eliminate external light, so we use a long shutter time (600 ms) to obtain sufficient exposure. A slight noise component is present in the images, but this is considered negligible. Finally, we use the estimated distortion coefficients to remove distortion from all images in the dataset so that subsequent algorithms may assume a pinhole camera model.

To avoid any compression or manipulation of the images by the camera software, in particular automatic color correction, we read the raw sensor data directly. We use bilinear interpolation to reconstruct RGB images from the raw Bayer pattern images. By doing this, we obtain a consistent RGB color space. Moreover, the raw sensor data are linear and correlate directly with radiometric quantities, which allows for better BRDF and environment map estimation in later stages of our pipeline.

We capture radiometrically relevant parts of our dataset in HDR by stacking multiple exposures [31]. More specifically, we stack 11 exposures at one-stop intervals ranging from 1 to 2048 ms. For the other parts of the dataset, we capture a single image at an exposure time of 600 ms.

B. Surface Reconstruction from Structured Light

We use a standard Gray code structured light approach to generate raw point clouds for a scene [32,33]. With camera parameters from the calibration, we transform these point clouds into the same world coordinate system.

To reconstruct one connected triangle mesh from the point clouds, we merge them into a single point cloud and perform screened Poisson reconstruction with trimming and an octree depth of nine [34]. This technique requires point normals, so before the merging, we generate normals for each point cloud as follows. We resample the point cloud down to 100,000 vertices via Poisson disk sampling [35] and then compute normals via planar fitting to a nearest neighborhood of 500 points (~16 mm radius). We then reorient all the normals according to the location of one of the cameras and transfer them back onto the original point cloud. This procedure ensures smooth continuous normals, necessary for a good performance of the mesh reconstruction algorithm. As we rely on smoothing, we cannot reconstruct features in the mesh with the same physical size as the alignment error accumulated from structured light and calibration. The aim of the chosen constants was to preserve features by striking a balance between too noisy and too smooth. The operability of the pipeline is, however, not sensitive to the choice of these constants.

C. Material BRDF Reconstruction

We assume that all non-transparent materials in the scene are opaque and isotropic, so we model their reflectance properties by BRDFs. To acquire a BRDF, we combine traditional canonical gonioreflectometric sampling [36] with a BRDF interpolation (reconstruction) technique [37]. We follow the workflow outlined in Fig. 3 (right). A light arc illuminates material samples from 11 unique inclinations, evenly distributed from 7.5° up to 90° with 7.5° steps. We place a flat material sample at the center of the circle partly traced by the light arc. Using the cameras mounted on the robot, we then measure radiance reflected by the sample across one octant of a

sphere. The center of this sphere coincides with that of the light arc, while its radius is slightly larger to avoid collision between the robot and the arc. The robot moves in steps of 7.5° and captures 11 HDR images of the sample per step, one for each light direction. In total, this yields 2,783 HDR images per material. We avoid tangential and zenith viewing directions (90° and 0°, respectively). In the former case, no reflected radiance should be visible, while in the latter the light arc occludes the view of the sample.

The 2,783 observations are too few to faithfully represent the BRDF of a material in a photorealistic rendering. We need an interpolation scheme to fill the entire $(90 \times 90 \times 180)$ Mitsubishi Electric Research Laboratories (MERL) format BRDF look-up table [38]. The reconstruction method by Nielsen *et al.* [37] is our interpolation scheme. First, we use each of the 100 BRDFs in the MERL dataset [38] as sample points in a $90 \cdot 90 \cdot 180 = 1,458,000$ dimensional space. The nonlinear mapping of Nielsen *et al.* [37] is then applied to each of the samples. The mapped samples are ordered as rows of a matrix $\mathbf{X} \in \mathbb{R}^{m \times d}$ where m is the number of BRDF samples, and d is the dimension of the space. The zero-mean matrix is computed as $\mathbf{X} - \bar{\mathbf{x}}$, with $\bar{\mathbf{x}}$ being the sample mean. From this, the singular value decomposition $\mathbf{X} - \bar{\mathbf{x}} = \mathbf{U\Sigma V}^T$ is used to compute the eigenvectors and eigenvalues of the covariance matrix of $\mathbf{X} - \bar{\mathbf{x}}$, which are given as the columns of $\mathbf{V}$ and the diagonal elements of Σ, respectively. This is effectively a principal component analysis (PCA), where the eigenvectors are the principal components. A matrix composed of the scaled principal components as columns are computed as $\mathbf{Q} = \mathbf{V\Sigma}$.

Now, the full BRDF can be reconstructed from this principal component space by projection. Let $x' \in \mathbb{R}^n$ be n BRDF observations measured for a given material. Then, let $\bar{\mathbf{x}}' \in \mathbb{R}^n$ be the mean values and $\mathbf{Q}' \in \mathbb{R}^{n \times k}$ be the scaled eigenvectors corresponding to the direction pairs of those n observations. A vector $\mathbf{c}$ that spans the full space can be constructed by finding the linear combinations of principal components that best approximate the n observations. We do this by solving the linear least-squares optimization problem given by

$$\mathbf{c} = \underset{\mathbf{c}}{\operatorname{argmin}} \| (\mathbf{x}' - \bar{\mathbf{x}})' - \mathbf{Q}'\mathbf{c} \|^2 + \eta \| \mathbf{c} \|^2$$

$$= (\mathbf{Q}'^T \mathbf{Q}' + \eta \mathbf{I})^{-1} \mathbf{Q}'^T (\mathbf{x}' - \bar{\mathbf{x}}').$$

Note that by adding a penalty η to the norm of $\mathbf{c}$, this effectively becomes a Tikhonov regularized least squares. Now, the full, mapped BRDF is reconstructed as $\mathbf{x} = \mathbf{Qc} + \bar{\mathbf{x}}$. The inverse of the nonlinear mapping applied to $\mathbf{X}$ is applied to $\mathbf{x}$ to get the actual, unmapped BRDF of the material. The described approach is applied to every single non-transparent material in the scene in order to obtain models of their reflectance properties.

This approach assumes that the MERL database encompasses the class of materials present in the scene. Effectively, this is a practical compromise between dense, unbiased, canonical BRDF sampling and fast, inferred BRDF sampling. This enables us to obtain high confidence BRDFs in a matter of a few hours.

D. Surface Reconstruction from CT

In our dataset, we have three glass objects: a sphere, a teapot (pot and lid), and a bowl (bowl and lid), for a total of five pieces. All objects have spherical plastic markers glued onto
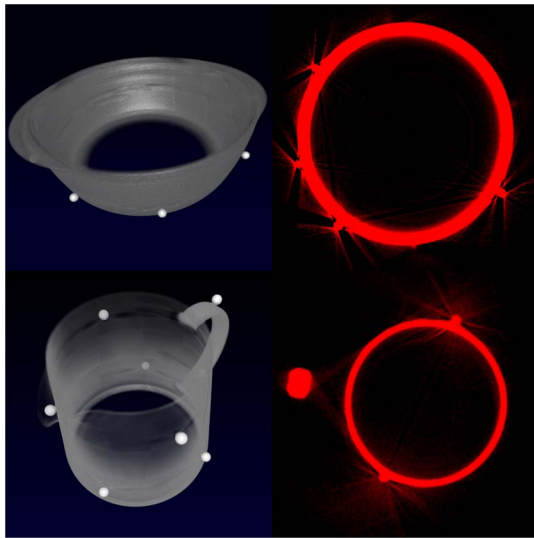
Fig. 4. CT scans of the bowl (top row) and the teapot (bottom row) with markers glued onto them. In the left column, visualized using a 1D transfer function. Note the different density of the markers. In the right column, a slice scaled to display streak artifacts.

Markers	After reconstruction	After simplification	After cleaning

Fig. 5. Reconstruction from CT with stages illustrated using Phong shading (top row) and wireframe shading (bottom row). After estimating the marker mesh (first column) and fitting spheres to the markers, we reconstruct the object mesh (second column). To eliminate noise, we first simplify the mesh (third column) and then close the holes and apply our subdivision-decimation loop to get the final object mesh (fourth column).

their outer surfaces. We CT scan each glass piece to obtain x-ray radiographs and use the CT PRO 3D reconstruction software from Nikon Metrology to obtain a volumetric image for each piece. The resolution of the reconstructed volume is up to 1000^3 voxels. Due to beam hardening, high-density differences between materials lead to streak artifacts [39], especially around our markers and at the tops and bottoms of the objects (see Fig. 4). We account for these artifacts in the volumetric segmentation.

From a CT scan, we generate two triangular meshes with vertex normals: one for the glass object and one the plastic markers. Figure 5 provides an overview of our procedure. We start with the markers, which appear as elements of higher density in the scan. We preprocess the scan by clamping all the values under a certain threshold to zero and then create a mesh using dual contouring [40]. Generating the glass mesh is more cumbersome. We also use dual contouring in this case, but because of the streak artifacts (Fig. 4), it is not possible to isolate the glass mesh via a threshold. Instead, we use a lower threshold that removes only noise, then estimate the marker positions and use these to remove the markers from the glass mesh.

To estimate marker positions, we determine a series of center/radius pairs $(\mathbf{c}_i, r_i)$ by fitting a multi-sphere model to the marker mesh vertices using a tuned random sample consensus (RANSAC) algorithm [41]. We then carve a hole by excluding all the triangles that are inside a sphere with center $\mathbf{c}_i$ and radius $(1 + \epsilon)r_i$, where ϵ is usually in the 0.5–0.75 range. We store the marker positions $\mathbf{c}_i$ so that we can use them to transform from the local coordinate system of the glass object to the world coordinate system (see Section 2.F).

After removing the markers, the glass meshes still have aliasing artifacts. To deal with this issue, we first decimate the mesh down to 1% of the original vertices via quadric edge collapse. The holes are then easy to close by identifying the edge loops surrounding each hole and filling these with triangles. We then
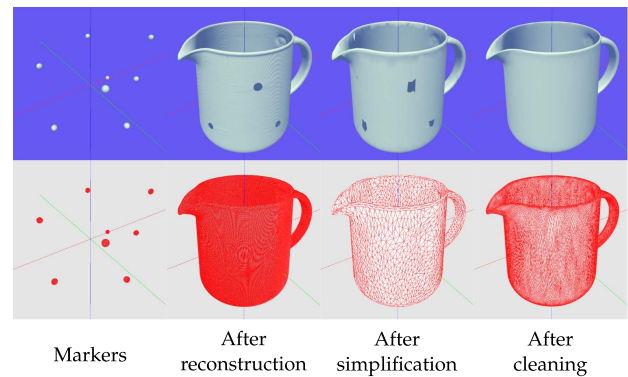
introduce a subdivision-decimation loop with alternating $\sqrt{3}$-subdivision [42] and decimation to 33% of the original vertices. We perform this subdivision-decimation operation four times to obtain a cleaned mesh. The decimation removes unwanted high-frequency features from the mesh. Thus, we generate smooth meshes at the cost of some geometric precision. We are again trying to strike a balance between reconstruction error and too much smoothing. In Section 4, we compare our method with a different cleaning procedure that better preserves geometry.

E. Scene Reassembly for Non-transparent Objects

Two operations are necessary to prepare the background mesh for rendering: labeling and deformation. In the labeling, our objective is to identify BRDFs and label each face of the mesh with a BRDF. Assuming a scene with a small number of known BRDFs, we apply edge detection and watershed on the images of the scene to segment BRDF boundaries. Shadows, specular highlights, and different viewing angles of the scene complicate fully automatic BRDF identification. Our approach gets us most of the way, but we manually correct any residual misclassification. Figure 6 shows a label image produced by our labeling technique.

The label images can be used in multi-view projective texturing of the background mesh. However, we would like to precompute the view and label selection instead of doing
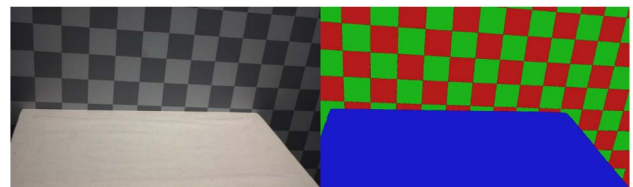

Fig. 6. Labeling of the image to the left results in the label image to the right. Each color in the label image represents a label that we assign a BRDF to. The black edges between labels indicate areas where we apply a nearest neighbor method.
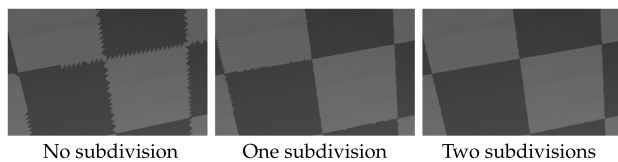
Fig. 7. Subdividing the mesh dissolves unwanted boundary saw-tooth artifacts that originate from the BRDF labeling.

it millions and millions of times while rendering. To avoid *uv*-unwrapping of the mesh for storing precomputed labels, we take an approach inspired by micropolygon rendering [43]. We project each vertex of a face onto the label images of the scene and select the face BRDF according to the image label that most of the face vertices were projected to. If a vertex projects to an unknown label, we resolve it by a nearest neighbor search. Since faces around material boundaries overlap multiple materials, we get sawtooth artifacts. We dissolve these by subdividing the mesh until the rendered triangles are smaller than the surface area observed in a pixel; see Fig. 7.

When applying physically based rendering, we observed intersections between background scene and glass meshes. This could be due to small errors in reconstruction and positioning, or perhaps the harder glass objects press down the tablecloth when placed for reference imaging. It causes significant visual artifacts, since the rendering exposes all surfaces of a transparent object. To eliminate these artifacts, we accommodate the hard object (glass) by deforming the soft object (tablecloth); see Fig. 8. To deform the soft object, we need a "down" direction in which to push the vertices. We first find contact vertices. These are vertices in each mesh that are close to any vertex of the other mesh. We consider vertices close if the distance between them is less than 7% of the bounding box diagonal of the hard object. Using least squares regression, we fit a contact plane to the contact vertices of the soft object. We set the sign of the contact plane normal so that the upper half-space contains the center of the hard object bounding box. Projection of a contact vertex to the normal of the contact plane then measures the height of the vertex. For each soft object contact vertex $\mathbf{x}$, we find the nearest hard object contact vertices and push $\mathbf{x}$ down below the lowest one of these.

F. Scene Reassembly for Transparent Objects

To reposition the glass objects in the scene, we rigidly transform the meshes reconstructed from CT to the world coordinate system of the background mesh. We obtain this transformation by matching markers in the stereo images with the marker coordinates $\mathbf{c}_i$ computed during reconstruction from CT (see Section 2.D).
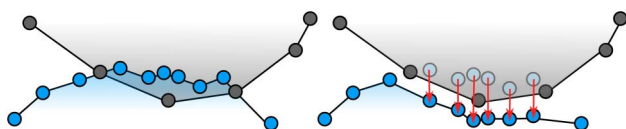

Fig. 8. Deformation of background mesh, where we push the background vertices down to avoid mesh intersection.
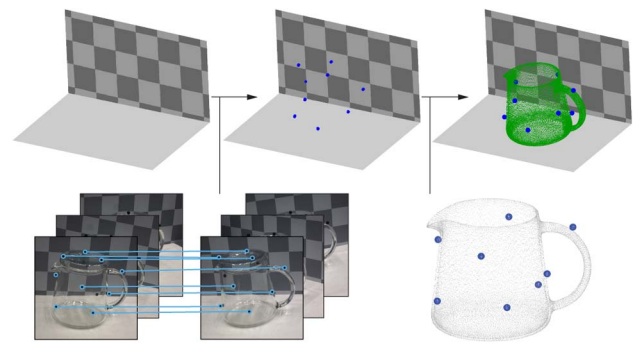

Fig. 9. Repositioning a CT scanned object in the background scene. We identify and match the markers in the stereo image pairs and calculate their corresponding 3D points. Pairing these with marker coordinates from the CT scans, we transform the CT scanned piece of an object into the world coordinate system.

To find the markers, we employ a size invariant circle Hough transform [44]. This works well for our dataset, where the markers show high contrast against their surroundings. We match markers in the left and the right images via Sampson distance [45]. Using this technique, markers on the same epipolar line lead to false positives, so we manually inspect the result. We also manually discard detected markers that are visible through the glass, as the refraction would lead to incorrect positioning. Markers in both stereo images with no match are discarded. The result is a set of matched markers in image coordinates, as seen in Fig. 9 (bottom left). We then triangulate the matched markers from the stereo views and gather them in clusters of 3D points. We remove outliers via their distances from the cluster centers, and for each cluster we select the point with the lowest reprojection error. An example of the points and clustering is shown in Fig. 9 (top middle).

We manually pair the 3D marker coordinates from the images with the marker coordinates $\mathbf{c}_i$ from the CT scans. We perform Procrustes analysis [46] on the two point sets, excluding reflection, since we assume a rigid transformation applied to each vertex of the mesh. The bowl and the teapot are composed of multiple pieces. For these objects, we compute the transformation individually for each piece. The result of the object transformed into the scene is shown in Fig. 9 (top right). We found that in order to have low error in the transformation, the chosen markers should sample the surface evenly and be visible from most views.

G. Color Calibration

Images are quantitatively comparable only if they live in the same color space. Thus, we must ensure that our radiometry-dependent data, namely reference images, environment map, and BRDFs, are in the same color space. We do this by imaging a color chart of precisely known colors. More specifically, we use second-degree root-polynomial color correction [47] based on a 24-patch ColorChecker Classic from X-Rite. This provides a matrix that transforms from camera RGB to XYZ, where we assume illuminant D50 when specifying the XYZ values of the colorchecker. With the assumption of illuminant D50, we can transform colors to the CIE L ∗ a ∗ b ∗ color
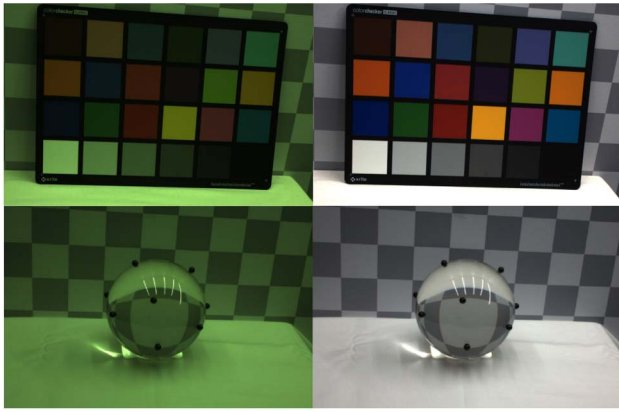
Fig. 10. Color calibration: raw images (left) and color corrected images (right). The camera sensor is particularly sensitive to green.

Fig. 11. Unwrapping of a spherical probe. We know the sphere radius R from specification, the camera position $\mathbf{c}$ through calibration, and the sphere center $\mathbf{o}$ by triangulation. Radiance at $\mathbf{p}_{\text{proj}}$ in our image then corresponds to the environment map direction $\vec{l}$. The result for the robot enclosure is in the lower left corner in latitude–longitude panoramic format (here tone-mapped).

space and then compute color differences using the ΔE_{00} metric [48]. We use this to refine our result by minimizing ΔE_{00} using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [49]. The result is in Fig. 10. The average color difference is $\Delta E_{00} = 1.97 \pm 1.21$, which is larger than 1 JND (just noticeable difference) [50], but we find it acceptable.

Since we work with glass objects (and chrome; see Section 2.H), we need refractive indices to determine reflectance, transmittance, and absorption properties. Refractive indices can be found per wavelength in tables of research papers. To use such spectral optical properties together with our trichromatic image data, we integrate them to CIE RGB using the CIE RGB color-matching functions listed by Stockman and Sharpe [51]. It is important to normalize these functions [52] and to use RGB rather than XYZ [53]. This is because a refractive index is not a color, but rather a quantity that in trichromatic representation should resemble a sparse sampling of the spectrum. Thus, as recommended by other authors [54], we choose CIE RGB as our rendering color space. After transforming our image data from camera RGB to XYZ, we therefore convert them to CIE RGB [55]. As a final step, we apply Bradford chromatic adaptation [50], adapting to the originally assumed illuminant D50, so that renderings and reference images get closer to real-life appearance.

H. Environment Lighting
To capture the lighting observed in the reference images, we use a method similar to the mirror probe technique [56]. However, we use a pinhole camera model for probe image unwrapping instead of the standard orthographic model. Our pipeline enables this, as we have a calibrated camera and know its position relative to the photographed mirror probe. With the pinhole model, we obtain a more precise estimate of the environment lighting. The environment map is generated from HDR images and stored in latitude–longitude panoramic format [50]. We use a polished grade G100 chrome bearing ball as mirror probe.

An environment map represents an infinite area light and maps a direction to a texture element (a texel). To do unwrapping, we map each texel d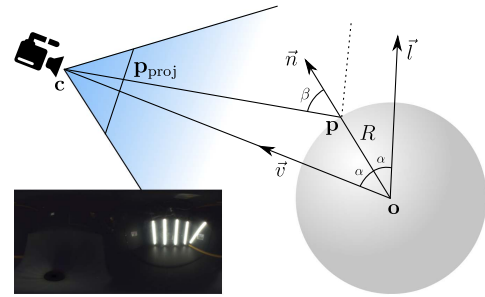irection $\vec{l}$ to the corresponding pixel position $\mathbf{p}_{\text{proj}}$ in a light probe image. Given the configuration illustrated in Fig. 11, we have

$$\vec{v} = \frac{\mathbf{c} - \mathbf{o}}{\|\mathbf{c} - \mathbf{o}\|}, \quad \vec{n} = \frac{\vec{v} + \vec{l}}{\|\vec{v} + \vec{l}\|}, \quad \mathbf{p} = \mathbf{o} + R\vec{n}, \quad \mathbf{p}_{\text{proj}} = \mathbf{M}[\mathbf{p}^T \quad 1]^T,$$

where camera matrix $\mathbf{M}$ and camera position $\mathbf{c}$ are available from our calibration. The radius of the sphere R is available from the bearing ball specification, and we find the center of the sphere $\mathbf{o}$ by manually annotating the sphere and then triangulating it. We assume that the distance to the actual light along $\vec{l}$ is equal to the distance between camera and sphere $\|\mathbf{c} - \mathbf{o}\|$. This assumption works well in practice, leading to an error smaller than the uncertainty of $\mathbf{o}$ caused by the triangulation. With the original orthographic camera model, we can reconstruct the lighting for all directions except one ($-\vec{v}$). In our model, we cannot reconstruct the lighting for a set of directions ($\vec{n} \cdot \vec{v} \leq R/\|\mathbf{c} - \mathbf{o}\|$), so we set them to black. Since we do our unwrapping in world space, we can combine contributions from multiple camera views with no need to align them afterwards.

The environment map is color corrected according to Section 2.G, which enables us to correct for the angularly dependent reflectance of chrome. The correction is to divide by Fresnel reflectance, which we compute during unwrapping. As input for Fresnel's equations, we use the angle β between $\mathbf{c} - \mathbf{p}$ and $\vec{n}$ and the complex refractive index of chrome [57] converted from spectrum to CIE RGB. The result is shown in the inset of Fig. 11.

I. Rendering
We render images using progressive unidirectional path tracing [58,59] implemented in OptiX [60]. The captured HDR environment map is the sole light source in our scene [56]. When rendering non-specular materials, we importance-sample the environment map to get direct illumination and use sampling of a cosine-weighted hemisphere to get indirect illumination. From our labeling, we have one BRDF attached to each triangle in our scene. For non-transparent objects, we use our measured BRDFs tabulated in the MERL format [38]. To terminate paths probabilistically, we use Russian roulette based on the

bihemispherical reflectance of each measured BRDF. This reflectance is calculated in a preprocessing step using Monte Carlo integration. We deal with transparent objects in the usual way, setting reflectance and transmittance according to Fresnel's equations of reflection and Bouguer's law of exponential attenuation. Given their small surface, we were unable to estimate a BRDF for the markers. Instead, we render them as glass with all refracted rays being absorbed.

3. ANALYSIS BY SYNTHESIS

The ability to render images comparable to photographs enables us to use our pipeline for improving parameter estimates through analysis by synthesis. As an example, we need a scaling factor for our HDR environment map, as it measures relative radiance [31]. We estimate this factor by taking ratios of references and renderings with the background scene alone. Another example is estimating real and imaginary parts of glass refractive indices. As analysis by synthesis is fundamentally ill posed [61], we take our outset in physics-based initial guesses such as Schott K5 crown glass (sphere and teapot) and soda lime glass (bowl). Spectral refractive indices for these glasses were obtained from an online database (http://refractiveindex.info) and converted to CIE RGB. All parameters were estimated using views different from the ones in our comparisons of renderings with references.

As an example of our analysis by synthesis, we plot the evolution of the root-mean-squared error (RMSE) for different renderings of the glass bowl in Fig. 12. For each rendering, we vary a trichromatic component of the absorption coefficient (which directly relates to the imaginary part of the refractive index). We identify a distinct minimum in the error for each channel, with a slightly larger uncertainty in the red channel. The minimum values in this figure were used in our renderings of the glass bowl. We apply the same analysis to the teapot and the sphere.

Given an initial guess for a parameter, we can employ standard optimization algorithms, defining the RMSE between the reference and the rendering as a cost function to minimize. To reduce rendering times, the evaluation of the cost function can be c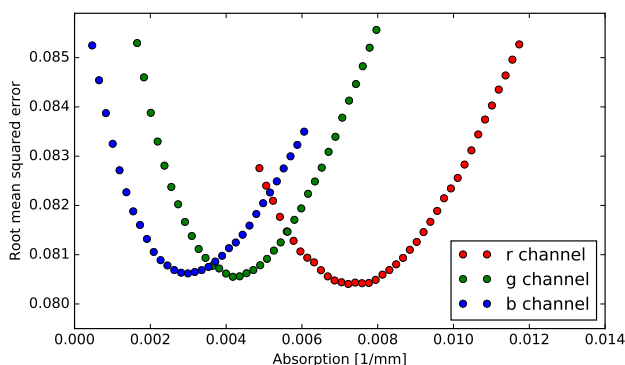alculated on a downsampled image or limited to a specific patch of the images. Various general optimization algorithms exist for minimizing expensive cost functions [62].

4. RESULTS

Our scenes consist of a backdrop, a stand, and a glass object (with markers) placed on the stand. The backdrop is a 30 by 20 white-and-gray checkerboard print on 120 cm by 80 cm semi-matte cardboard, and the stand is a tabletop with a white cloth. An example scene is depicted in Fig. 13. We implemented our reconstruction and reassembly procedures as a modular software pipeline and computed all rendered images using our path tracer. As illustrated in Fig. 2 and mentioned in Section 2.G, we color correct both rendered images and reference images to have a meaningful perceptual comparison. Figure 14 compares markers in a reference image with rendered markers to validate our marker positioning. For the teapot, the average distance between the markers from stereo and the transformed markers from CT is 0.43 mm.

Figure 15 presents pixelwise comparisons of reference images and rendered images. The error images allow us to spot subtle differences not easily noticed in a perceptual comparison, such as the slight misalignments in geometry and highlights. As reference photographs were not captured in HDR, we clamp the renderings correspondingly. This means that areas of strong light intensity, such as highlights and intense caustics, appear black in the error images.

Fig. 13. Scene with checkerboard backdrop, lighting, glass teapot, and stand with table cloth observed by two cameras mounted on a 6-axis industrial robot arm.

Fig. 12. Analysis by synthesis to estimate absorption of the glass bowl. We run renderings in low resolution and change the absorption in each color channel one at the time. In the case of the bowl, the blue channel is the most sensitive one.

Fig. 14. Markers rendered in blue and added to the reference image to validate marker positions by looking at pixel offsets.
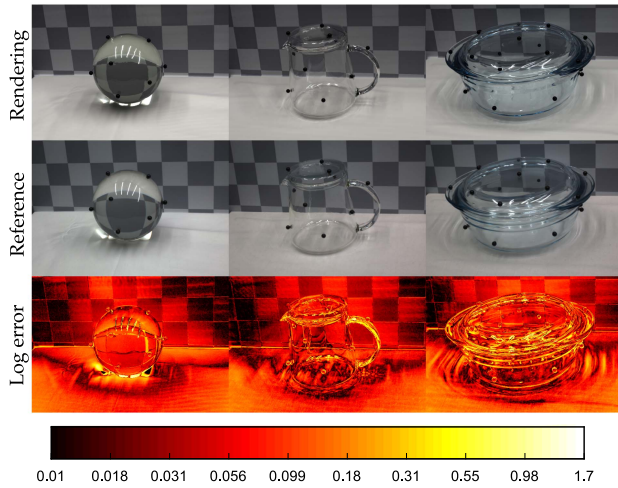
Fig. 15. Pixelwise error for three rendering-reference pairs. Error is the ℓ^2-norm of 32-bit per channel RGB images, visualized using a base 10 logarithmic scale.

Fig. 17. Zoom-in of Figs. 16(b) and 16(c) to emphasize the effect of our background deformation.

Orthographic   Perspective   Reference

Fig. 18. Zoom-in of Fig. 16(c) and 16(d) to emphasize the effect of our perspective unwrapping of the environment map.

Figure 16 exemplifies the impact on error images of some of our contributions. In Fig. 16(a), we reposition only the glass object in the background scene and apply color correction (Sections 2.F and 2.G). This means that we use Lambertian materials (with bihemispherical reflectances from the measured BRDFs), an orthographic unwrapping model of the environment map, and no chrome reflectance correction or analysis by synthesis optimization. We compare to the reference image in Fig. 16(g), with error images as in Fig. 15. Figure 16(b) shows the impact of using measured BRDFs (Section 2.C), resulting in a more accurate representation of the folds of the cloth in the background scene (top image) and an overall reduction of the error (bottom image). In Fig. 16(c), we add deformation of the background mesh (Section 2.E), which ensures that the background mesh does not poke through the glass surface (see a close-up in Fig. 17). Additionally, we can see how this improves the error on the lid of the bowl, because of refraction of light in the glass. The next step, Fig. 16(d), shows the impact of our modified environment map unwrapping (Section 2.H) against the standard orthographic unwrapping
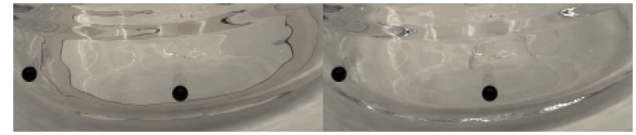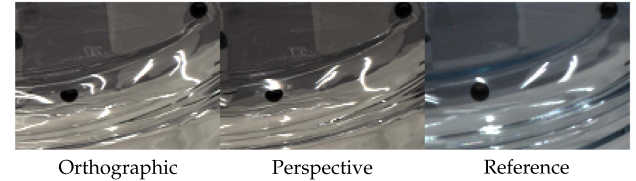
rotated according to our camera parameters. A close-up is available in Fig. 18. Our modified unwrapping provides a better shape and alignment of highlights and caustics. Partially due to the assumption of infinitely distant environment light, some alignment artifacts persist. In Fig. 16(e), we show the effect of correcting for chrome reflectance in our environment map reconstruction. Quantitatively, this changes the distribution of the error (bottom image). On the cloth, the exposure increases, exposing the caustics misalignment. On the backdrop, the error reduces. Interestingly, the structural similarity index (SSIM) improves while the RMSE worsens. Finally, in Fig. 16(f), we use analysis by synthesis to adjust glass absorption. This improves the glass appearance, but it also leads to slight color changes in other parts of the scene due to indirect light paths. Because of this global influence, the analysis by synthesis introduces slightly too much absorption to compensate for the slightly too bright tablecloth.

As an example of how our pipeline can be used to validate existing algorithms, we investigate the case of glass object reconstruction. In Fig. 19, we compare two different reconstruction
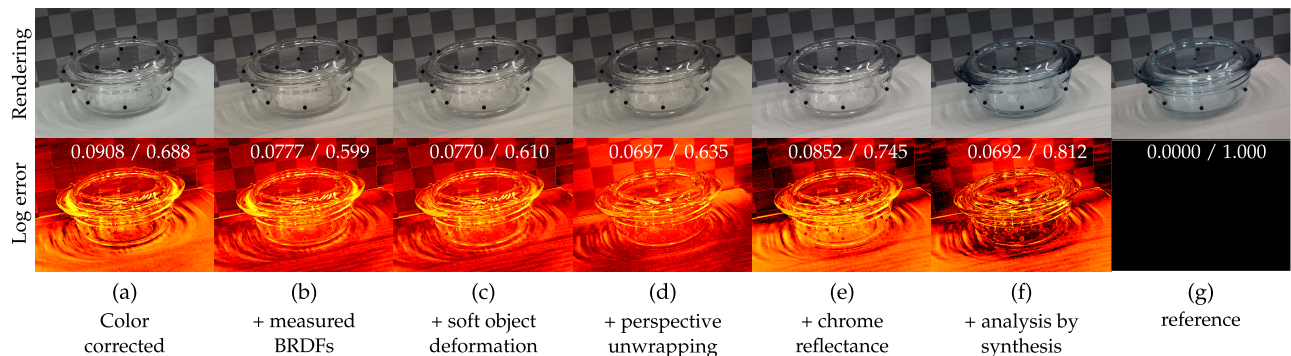

(a)	(b)	(c)	(d)	(e)	(f)	(g)
Color corrected	+ measured BRDFs	+ soft object deformation	+ perspective unwrapping	+ chrome reflectance	+ analysis by synthesis	reference

Fig. 16. Qualitative (top) and quantitative (bottom) step-by-step evaluation of our reassembly techniques. The log error images have the same format as in Fig. 15 and the reference photograph is in the rightmost column (g). In each column, we provide root-mean-squared error and structural similarity index (RMSE/SSIM). Both measures attain their best scores in our final result (f).
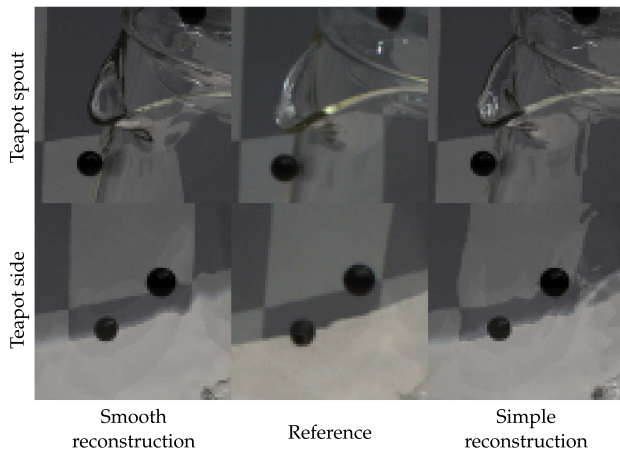
Teapot spout

Teapot side

Smooth reconstruction   Reference   Simple reconstruction

Fig. 19.   Trade-off in mesh reconstruction. If we smooth more, we get less distortion in the refractions, but less precision in the mesh geometry. From left to right: Rendering with smoothing, reference image, rendering without smoothing.

methods with a focus on two parts of the teapot scene. Smooth reconstruction refers to the procedure described in Section 2.D. The other procedure is simply to decimate the reconstructed mesh to 2.5% of the original vertices and apply Taubin smoothing [63]. This removes the high frequencies of the noise, but much noise is still present in the midranges, leading to wobbly refractions. Our method in Section 2.D reduces far more noise, but this is at the cost of greater changes to the overall shape. We note that a refractive object with a simple geometry is very hard to reconstruct automatically if fidelity and almost no noise are both required.

5. DISCUSSION

Since our pipeline enables us to compare renderings with photographs, we can identify problems in acquisition, reconstruction, and rendering that would otherwise have been hard to find. Camera calibration issues, for example, reveal themselves as error lines along edges (visible in Fig. 20). Color calibration issues reveal themselves as color shift. Such issues led us to more careful camera calibration procedures and the choice of root-polynomial color correction. Qualitative comparisons revealed artifacts in surface reconstruction, mesh intersections calling for deformation, misplacement of highlights, color shift due to chrome reflectance, and missing absorption in renderings (Figs. 16–19). Quantitative comparisons confirmed
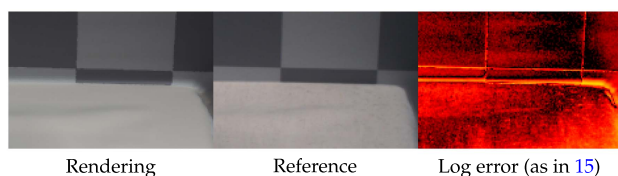

Rendering   Reference   Log error (as in 15)

Fig. 20.   Material transitions: error lines along checker edges and along the boundary between tablecloth and backdrop.

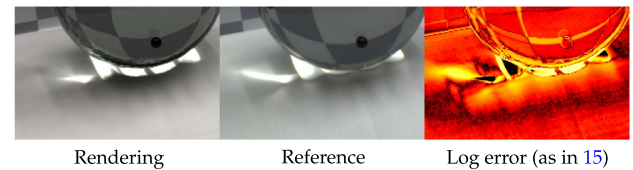Rendering   Reference   Log error (as in 15)

Fig. 21.   Effect of separating markers from glass (refracted light close to marker) and of not accounting for subsurface scattering (dark areas close to caustics).

improvement due to perspective unwrapping of light probe images and led to analysis by synthesis.

The comparison with reference photographs before and after deformation (Fig. 17) to some extent validates our soft object deformation technique. Further validation would be desirable, but it is difficult to come up with a different experiment. Some kind of soft, durable memory foam with a scannable surface would be required, as the soft object would otherwise change shape again once the hard object is removed. Our validation supports only that the cloth appearance (as observed through glass) is represented more faithfully after deformation.

We found analysis by synthesis useful for estimating parameters with an outset in physics-based initial guesses. The results in Fig. 12 show that we can estimate optical properties for a given material and use them in a different setting (right part of Fig. 1). The precision of the estimation varies with the impact of the property on the overall error, and the estimated parameters may compensate for unrelated errors. In this regard, specific scene configurations could be used to favor estimation of a particular parameter. The most important limitation of our method is that we describe materials as large patches of isotropic BRDFs. In our renderings, this assumption works well for the checkerboard backdrop but not for the cloth, where we have both subsurface scattering effects and probably anisotropy due to the weave structure of the cloth. Figure 21 reveals that the rendered image is too dark in areas surrounding caustics. As seen in the light refracted through the sphere in the vicinity of the marker, our processing of the glass object to separate glass from markers causes some imprecision in the geometry. We believe this mainly influences the shape of the caustic. The bleeding of the caustic to areas that are much darker in the rendered images looks like backscattering from the table beneath the cloth. We refer to this as a kind of subsurface scattering.

Another limitation is seen at the transition between non-connected elements. It is visible in the renderings at the boundary between the cloth and the backdrop (see Fig. 20). The problem derives from the fact that the cloth and the backdrop were too close to each other during dataset acquisition. This resulted in the Poisson mesh reconstruction interpreting them as a continuous object instead of two separate ones. The problems around markers (Fig. 21) are also due to transition of materials. The marker removal and whole closing in the glass surface reconstruction interrupts the original shape of the surface. Furthermore, the markers are glued onto the glass surface, and the glue is not considered in the reconstruction and renderings. The marker glue problem is magnified by the glass refraction.

6. CONCLUSION

We have proposed a pipeline for multimodal scene digitization. Our work addresses the entire process from acquisition of the original objects, through reassembly of the digital scene, to accurate modeling of camera and environment. While the pipeline required several non-trivial steps, the benefits are correspondingly great, since we can perform pixelwise comparisons between rendered images and photographs of the corresponding physical scene. This means that we have the means to quantitatively assess the accuracy of an acquired model based on comparison with empirical evidence. We believe this kind of quantitative assessment has not previously been possible for transparent objects. In applications such as cultural heritage preservation and industrial inspection, where the accuracy of a digitization is important, such comparison with empirical evidence is crucial.

To the best of our knowledge, our work is also the first work to quantify the photorealism of a heterogeneous scene requiring multimodal acquisition.

Our dataset is publicly available so that others can test new techniques for the different steps of the pipeline with quantitative feedback based on photorealistic rendering. The fact that one can use off-the-shelf rendering techniques for improving the different steps of a multimodal digitization pipeline is perhaps the most important benefit of our work. An application of the full pipeline is the virtual product placement in Fig. 1. Another important application is the estimation of radiometric properties through analysis by synthesis. The ability to accurately estimate optical properties through computation rather than measurement, which might require specialized equipment, is likely to greatly simplify the digitization of radiometrically complex objects. In this paper, we estimated absorption and refractive indices of transparent objects, but analysis by synthesis could be equally useful for other materials with non-trivial BRDFs. This is another key benefit of our work that we believe is well worth exploring in the future.
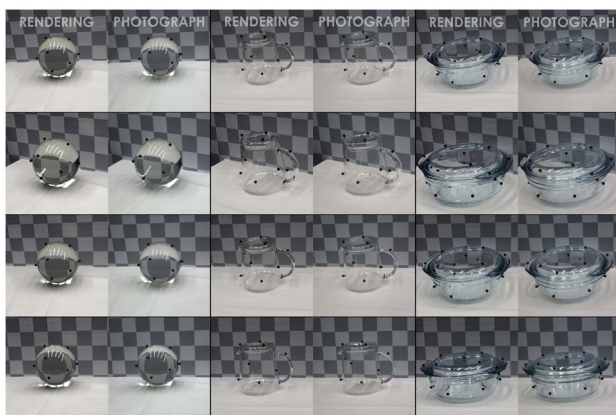
APPENDIX A

Figure 22.

Fig. 22. Comparison of renderings and photographs as in Fig. 1 (left), but with more views.

REFERENCES

1. M. Weinmann and R. Klein, "Advances in geometry and reflectance acquisition (course notes)," in *SIGGRAPH Asia 2015 Courses* (ACM, 2015).
2. P. Debevec, "The light stages and their applications to photoreal digitial actors," in *SIGGRAPH Asia 2012 Technical Briefs* (2012).
3. L. Gomes, O. R. P. Bellon, and L. Silva, "3D reconstruction methods for digital preservation of cultural heritage: a survey," Pattern Recogn. Lett. **50**, 3–14 (2014).
4. L. Zhang, H. Dong, and A. E. Saddik, "From 3D sensing to printing: a survey," ACM Trans. Multimedia Comput. Commun. Appl. **12**, 27 (2016).
5. J. B. Nielsen, E. R. Eiriksson, R. L. Kristensen, J. Wilm, J. R. Frisvad, K. Conradsen, and H. Aanæs, "Quality assurance based on descriptive and parsimonious appearance models," in *Workshop on Material Appearance Modeling (MAM)* (The Eurographics Association, 2015), pp. 21–24.
6. B. T. Phong, "Illumination for computer generated pictures," Commun. ACM **18**, 311–317 (1975).
7. C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile, "Modeling the interaction of light between diffuse surfaces," SIGGRAPH Comput. Graph. **18**, 213–222 (1984).
8. A. Takagi, H. Takaoka, T. Oshima, and Y. Ogata, "Accurate rendering technique based on colorimetric conception," SIGGRAPH Comput. Graph. **24**, 263–272 (1990).
9. G. W. Meyer, H. E. Rushmeier, M. F. Cohen, D. P. Greenberg, and K. E. Torrance, "An experimental evaluation of computer graphics imagery," ACM Trans. Graph. **5**, 30–50 (1986).
10. H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust, "Comparing real and synthetic images: some ideas about metrics," in *Conference: Proceedings of the Eurographics Workshop Rendering Techniques* (Springer, 1995), pp. 82–91.
11. K. F. Karner and M. Prantl, "A concept for evaluating the accuracy of computer generated images," in *Spring Conference on Computer Graphics (SCCG)* (1996).
12. S. N. Pattanaik, J. A. Ferwerda, K. E. Torrance, and D. P. Greenberg, "Validation of global illumination solutions through CCD camera measurements," in *Color Imaging Conference (CIC)* (1997), pp. 250–253.
13. N. L. Jones and C. F. Reinhart, "Parallel multiple-bounce irradiance caching," Comput. Graph. Forum **35**, 57–66 (2016).
14. N. L. Jones and C. F. Reinhart, "Experimental validation of ray tracing as a means of image-based visual discomfort prediction," Build. Environ. **113**, 131–150 (2017).
15. D. P. Greenberg, K. E. Torrance, P. Shirley, J. Arvo, J. A. Ferwerda, S. Pattanaik, E. Lafortune, B. Walter, S.-C. Foo, and B. Trumbore, "A framework for realistic image synthesis," in *SIGGRAPH 97* (ACM/ Addison-Wesley, 1997), pp. 477–494.
16. F. Drago and K. Myszkowski, "Validation proposal for global illumination and rendering techniques," Comput. Graph. **25**, 511–518 (2001).
17. C. Ulbricht, A. Wilkie, and W. Purgathofer, "Verification of physically based rendering algorithms," Comput. Graph. Forum **25**, 237–255 (2006).
18. J. Meseth, G. Müller, R. Klein, F. Röder, and M. Arnold, "Verification of rendering quality from measured BTFs," in *Applied Perception in Graphics and Visualization (APGV)* (ACM, 2006), pp. 127–134.
19. A. I. Ruppertsberg and M. Bloj, "Rendering complex scenes for psychophysics using RADIANCE: how accurate can you get?" J. Opt. Soc. Am. A **23**, 759–768 (2006).
20. A. Dal Corso, J. R. Frisvad, T. K. Kjeldsen, and J. A. Bærentzen, "Interactive appearance prediction for cloudy beverages," in *Workshop on Material Appearance Modeling (MAM)* (The Eurographics Association, 2016), pp. 1–4.

21. 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 Trans. Graph. **32**, 109 (2013).

22. T. Nöll, J. Köhler, G. Reis, and D. Stricker, "Fully automatic, omnidirectional acquisition of geometry and appearance in the context of cultural heritage preservation," J. Comput. Cult. Herit. **8**, 2 (2015).

23. H. Wu, Z. Wang, and K. Zhou, "Simultaneous localization and appearance estimation with a consumer RGB-D camera," IEEE Trans. Vis. Comput. Graph. **22**, 2012–2023 (2016).

24. I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich, "Transparent and specular object reconstruction," Comput. Graph. Forum **29**, 2400–2426 (2010).

25. A. Kolb, J. Zhu, and R. Yang, "Sensor fusion," in *Digital Representation of the Real World*, M. A. Magnor, O. Grau, O. Sorkine-Hornung, and C. Theobalt, eds. (CRC Press, 2015), Chap. 9, pp. 133–150.

26. V. Bhateja, H. Patel, A. Krishn, A. Sahu, and A. Lay-Ekuakille, "Multimodal medical image sensor fusion framework using cascade of wavelet and contourlet transform domains," IEEE Sens. J. **15**, 6783–6790 (2015).

27. A. Pamart, O. Guillon, J.-M. Vallet, and L. De Luca, "Toward a multimodal photogrammetric acquisition and processing methodology for monitoring conservation and restoration studies," in *Eurographics Workshop on Graphics and Cultural Heritage* (The Eurographics Association, 2016), pp. 207–210.

28. H. Aanæs and A. B. Dahl, "Accuracy in robot generated image data sets," in *Scandinavian Conference on Image Analysis (SCIA)*, Lecture Notes in Computer Science (Springer, 2015), Vol. **9127**, pp. 472–479.

29. H. Aanæs, R. R. Jensen, G. Vogiatzis, E. Tola, and A. B. Dahl, "Large-scale data for multiple-view stereopsis," Int. J. Comput. Vis. **120**, 153–168 (2016).

30. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000).

31. P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *SIGGRAPH 97* (ACM/Addison-Wesley, 1997), pp. 369–378.

32. J. L. Posdamer and M. Altschuler, "Surface measurement by space-encoded projected beam systems," Comput. Graph. Image Process. **18**, 1–17 (1982).

33. J. Geng, "Structured-light 3D surface imaging: a tutorial," Adv. Opt. Photon. **3**, 128–160 (2011).

34. M. Kazhdan and H. Hoppe, "Screened Poisson surface reconstruction," ACM Trans. Graph. **32**, 29 (2013).

35. M. Corsini, P. Cignoni, and R. Scopigno, "Efficient and flexible sampling with blue noise properties of triangular meshes," IEEE Trans. Visualization Comput. Graph. **18**, 914–924 (2012).

36. J. F. Murray-Coleman and A. M. Smith, "The automated measurement of BRDFs and their application to luminaire modeling," J. Illum. Eng. Soc. **19**, 87–99 (1990).

37. J. B. Nielsen, H. W. Jensen, and R. Ramamoorthi, "On optimal, minimal BRDF sampling for reflectance acquisition," ACM Trans. Graph. **34**, 186 (2015).

38. W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," ACM Trans. Graph. **22**, 759–769 (2003).

39. J. F. Barrett and N. Keat, "Artifacts in CT: recognition and avoidance," RadioGraphics **24**, 1679–1691 (2004).

40. T. Ju, F. Losasso, S. Schaefer, and J. Warren, "Dual contouring of Hermite data," ACM Trans. Graph. **21**, 339–346 (2002).

41. M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM **24**, 381–395 (1981).

42. L. Kobbelt, "√3-subdivision," in *SIGGRAPH 2000* (ACM/Addison-Wesley, 2000), pp. 103–112.

43. R. L. Cook, "The Reyes image rendering architecture," ACM SIGGRAPH Comput. Graph. **21**, 95–102 (1987).

44. T. Atherton and D. Kerbyson, "Size invariant circle detection," Image Vis. Comput. **17**, 795–803 (1999).

45. P. D. Sampson, "Fitting conic sections to "very scattered" data: an iterative refinement of the Bookstein algorithm," Comput. Graph. Image Process. **18**, 97–108 (1982).

46. J. C. Gower, "Generalized procrustes analysis," Psychometrika **40**, 33–51 (1975).

47. G. D. Finlayson, M. Mackiewicz, and A. Hurlbert, "Color correction using root-polynomial regression," IEEE Trans. Image Process. **24**, 1460–1470 (2015).

48. G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations," Color Res. Appl. **30**, 21–30 (2005).

49. J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, 2006).

50. E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*, 2nd ed. (Morgan Kaufmann/Elsevier, 2010).

51. A. Stockman and L. T. Sharpe, "The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype," Vis. Res. **40**, 1711–1737 (2000).

52. J. R. Frisvad, N. J. Christensen, and H. W. Jensen, "Computing the scattering properties of participating media using Lorenz-Mie theory," ACM Trans. Graph. **26**, 60 (2007).

53. C. Ulbricht and A. Wilkie, "A problem with the use of XYZ colour space for photorealistic rendering computations," in *Colour in Graphics, Imaging, and Vision (CGIV)* (2006), pp. 435–437.

54. J. Meng, F. Simon, J. Hanika, and C. Dachsbacher, "Physically meaningful rendering using tristimulus colours," Comput. Graph. Forum **34**, 31–40 (2015).

55. H. S. Fairman, M. H. Brill, and H. Hemmendinger, "How the CIE 1931 color-matching functions were derived from Wright-Guild data," Color Res. Appl. **22**, 11–23 (1997).

56. P. Debevec, "Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *SIGGRAPH 98* (ACM, 1998), pp. 189–198.

57. A. D. Rakić, A. B. Djurišić, J. M. Elazar, and M. L. Majewski, "Optical properties of metallic films for vertical-cavity optoelectronic devices," Appl. Opt. **37**, 5271–5283 (1998).

58. J. T. Kajiya, "The rendering equation," ACM SIGGRAPH Comput. Graph. **20**, 143–150 (1986).

59. M. Pharr, W. Jakob, and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*, 3rd ed. (Morgan Kaufmann/Elsevier, 2017).

60. S. G. Parker, J. Bigler, A. Dietrich, H. Friedrich, J. Hoberock, D. Luebke, D. McAllister, M. McGuire, K. Morley, A. Robison, and M. Stich, "OptiX: a general purpose ray tracing engine," ACM Trans. Graph. **29**, 66 (2010).

61. M. Hejrati and D. Ramanan, "Analysis by synthesis: 3D object recognition by object reconstruction," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2014), pp. 2449–2456.

62. D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," J. Global Optim. **13**, 455–492 (1998).

63. G. Taubin, "A signal processing approach to fair surface design," in *SIGGRAPH 95* (ACM, 1995), pp. 351–358.