Multi-View Normal Field Integration for 3D Reconstruction of Mirroring Objects

Michael Weinmann, Aljosa Osep, Roland Ruiters, Reinhard Klein
University of Bonn
Friedrich-Ebert-Allee 144, 53113 Bonn
mw@cs.uni-bonn.de, osep@cs.uni-bonn.de, ruiters@cs.uni-bonn.de, rk@cs.uni-bonn.de

Abstract

In this paper, we present a novel, robust multi-view normal field integration technique for reconstructing the full 3D shape of mirroring objects. We employ a turntable-based setup with several cameras and displays. These are used to display illumination patterns which are reflected by the object surface. The pattern information observed in the cameras enables the calculation of individual volumetric normal fields for each combination of camera, display and turntable angle. As the pattern information might be blurred depending on the surface curvature or due to non-perfect mirroring surface characteristics, we locally adapt the decoding to the finest still resolvable pattern resolution. In complex real-world scenarios, the normal fields contain regions without observations due to occlusions and outliers due to interreflections and noise. Therefore, a robust reconstruction using only normal information is challenging. Via a non-parametric clustering of normal hypotheses derived for each point in the scene, we obtain both the most likely local surface normal and a local surface consistency estimate. This information is utilized in an iterative min-cut based variational approach to reconstruct the surface geometry.

1. Introduction

3D reconstruction is one of the fundamental problems in computer vision. It has remained in the focus of research since decades with many applications in *e.g.* industry, entertainment and cultural heritage. While a huge amount of techniques has been developed in this field, today's challenges can be found when considering surfaces which exhibit a complex surface reflectance behavior. In this paper, we focus on reconstructing mirroring objects. For such objects, most traditional techniques such as laser scanners, structured light or multi-view stereo are not applicable.

Assuming a perfect mirroring surface, the appearance of a surface point only depends on the surrounding environment, the viewing angle and the local surface normal. By controlling the environment, it is directly possible to estimate normal information [29, 9, 16]. An alternative is to rotate the object and to track the optical flow [1, 30].

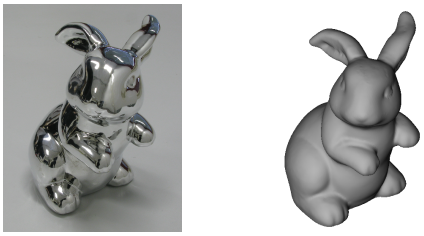

Figure 1: Bunny figurine and reconstructed model.

Several approaches such as the ones in [9, 16] use these normals to perform a single-view normal field integration and are thus limited to partial 2.5D reconstructions. Others derive a normal consistency measure and perform a multi-view reconstruction (*e.g.* [6, 27]). However, normal consistency alone is not suitable to reconstruct fine surface details. Therefore, a final refinement step is performed in [27] to combine the geometry estimated from the normal consistency with the observed surface normals. However, none of the mentioned approaches has shown high-quality reconstructions for complex geometries in the presence of occlusions and interreflections.

To address this problem, we exploit the fact that outliers due to occlusions or interreflections are not consistent for different measurements taken under varying configurations of viewpoint and light source position. Inspired by the multi-view normal field integration approach presented in [8] but utilizing a numerical scheme for obtaining a globally consistent surface reconstruction similar to [35], we formulate the problem in terms of an optimization which combines both a local surface consistency measure

and the observed normal information. We determine these quantities in an outlier-robust way via mean-shift clustering [10] of the individual local normal hypotheses which result from different configurations of viewpoint and light position. This makes our approach capable of handling occlusions. To acquire the full shape of the considered object, the utilized setup comprises a turntable, eleven cameras and three screens for displaying structured light patterns which are reflected by the object surface. Our technique produces high-quality reconstructions of the full 3D shape of an object not only on synthetic but also on real-world data (see *e.g*. Figure 1).

In summary, the key contributions of our approach include a system for acquiring the full 3D shape of mirroring objects based on multi-view normal field integration and a novel clustering-based scheme for integrating different volumetric normal fields which is robust in the presence of outliers and noise and makes accurate 3D reconstruction possible on real-world data.

2. Related Work

Surface reconstruction has attracted a lot of research in the last decades. We focus on giving a brief review on normal-based reconstruction techniques and approaches for 3D reconstruction of highly specular and mirroring objects.

Amongst the early investigations for exploring normal information for 3D reconstruction are shape-from-shading techniques [20] and photometric stereo [36] which focused on reconstructing Lambertian objects from a single view under known light source positions. Since then, many techniques focused on extending photometric stereo towards general unknown illumination [4, 37] and providing robustness to violations of the underlying assumption of Lambertian reflectance behavior due to specularities or shadows. Other methods addressed a more general surface reflectance behavior such as spatially-varying BRDFs [18, 17]. However, effects such as shadows or interreflections are not taken into account. Furthermore, multi-view photometric stereo has been explored in *e.g*. [14, 5]. Targeting on the larger range of opaque materials, a reciprocal setup where camera and light source positions can be exchanged in order to exploit the Helmholtz reciprocity for calculating surface normals has been proposed in [42]. This principle has further been investigated in [13] in a multi-view setting.

Focusing on the reconstruction of specular objects, we refer to the surveys in [21, 3] and the theoretical discussion in [23]. Methods such as specular flow techniques [28, 1] compute the surface geometry from the movement of the environment features mirrored on the object surface. Such methods usually rely on a known motion of the mirroring object, its environment or the cameras respectively. However, estimating dense optical flow is non-trivial due to the possibility of observing a single environment feature several times on the specular surface due to interreflections and usually a distant environment is assumed. For this reason, sparse reflectance correspondences have been used to locally approximate specular surfaces using quadrics in [30].

Other reconstruction approaches investigate the use of specular highlights observed on the object surface due to specular reflection in controlled environments. For this purpose, it is required to obtain dense observations of such specularities on the specular surface. This can be performed by moving the camera [43], using a moving light source [9], using extended light sources [22] or sequentially switching on individual elements of a grid of light sources [29]. As the number of required images increases linearly with the utilized light source positions, some techniques aim at significantly reducing the amount of required images by performing measurements in parallel. This can be achieved by rotating the object and using a circular light source [41] or by using printed, static or moving calibrated patterns [6, 31, 24]. Furthermore, some methods encode multiple light sources simultaneously. While encoding schemes for light source arrays have been investigated in [26], several approaches extend this idea by simulating a dense illumination arrays using LCD screens and encode the illumination emitted from the pixels using structured light patterns [16, 27, 38, 2]. However, in most of the approaches the assumption of far-field illumination or a distant environment is violated as the LCD display or the printed patterns have to be placed closely to the object for obtaining a sufficient sampling of light directions or feature directions.

The resulting normal-depth ambiguity can be solved in a multi-view setting such as the one presented in [6] where a calibrated pattern is used to produce reflections on the specular surface. Based on a volumetric representation, the law of reflection is used to hypothesize a normal at each voxel. Subsequently, the surface is assumed to pass through the voxels with the most consistent normal hypotheses following a normal disparity measure. The idea of hypothesizing surface normals has already been investigated in *e.g*. [11, 25] for extending classical single-view photometric stereo by selecting only hypotheses which agree with the underlying model assumptions. In [19], several normal hypotheses are generated for each pixel from different lighting directions. The solution space is then reduced according to an agreement concerning monotonicity, visibility, and isotropy properties. This makes the approach applicable for both diffuse and specular surfaces. Instead of a reduced solution space, the approach presented in [5] determines a maximal set of inliers per voxel on which regular photometric stereo is applied in a multi-view approach. While producing good reconstructions on synthetic data, the estimated surface consistency tends to being localized non-accurately for real-world data due to the lack of a per-voxel normalization. Further investigations on matching

hypothesized normal information in the context of specular surface reconstruction include the approaches proposed in [34, 2]. In [2], overlapping deflectometric measurements from multiple views are used to reconstruct large mirroring surfaces. However, self-occlusions represent problems for this approach and the configuration of the individual views has to be performed manually. Clustering normal observations per pixel in a single-view setting via the k-means algorithm has been used in [39] for reconstructing transparent objects. Similar to [6], specularity consistency between a set of views in a triangulation-based scheme using a display with Gray codes for illumination has been investigated in [27]. After triangulation, normals are refined for the estimated depth values in a similar way to the iterative scheme proposed in [32].

Closely related to our approach are the multi-view normal field integration approaches proposed in [8] and [12] in the context of photometric stereo. These overcome the problem of obtaining only 2.5D reconstructions of partial surfaces in the single-view case. In [8], an initial visual hull reconstruction is followed by an iterative surface evolution based on level sets in a variational formulation. As no global optimization is performed, the surface evolution is sensitive to the initial visual hull. In contrast, the technique proposed in [12] is based on a Markov Random Field (MRF) energy function where the surface is computed via min-cut to find a global minimum. This is followed by a smoothing step similar to the one applied in [8]. A surface orientation constraint has been included in the energy functional which enforces the reconstructed geometry to agree with the observed surface normals. Both techniques employ additional silhouette information which is very difficult to determine for mirroring objects. In contrast, our method only incorporates normal information and brings the multi-view normal integration to the domain of reconstructing mirroring objects.

3. Problem Statement

Given a set of κ_c calibrated cameras $\mathcal{C}_i$ with $i = 1, \ldots, \kappa_c$ which are positioned to observe a mirroring object from different viewpoints and a set of $j = 1, \ldots, \kappa_s$ screens, our goal is to reconstruct the object surface δV of a mirroring object with volume V by utilizing only normal information recovered for the individual views. Apart from a smoothness prior, we do not incorporate any prior knowledge about the object geometry such as the assumption of rather flat surfaces [9, 16] or an initial visual hull reconstruction [8]. Furthermore, our approach should consider the possibility of self-occlusions of the object geometry. Due to the complexity of real-world scenarios, we also have to design our reconstruction technique to be robust to noise. In addition, violations of the assumption regarding the underlying reflectance model need to be handled to some degree as well as incomplete normal fields which occur when no normal information can be derived for certain parts of the object surface. For this reason, we formulate the surface reconstruction as a variational energy minimization problem similar to [8] according to

$$\min_V \left\{ -\lambda_1 \int_{\delta V} \langle c\mathbf{N}, \mathbf{n} \rangle \, \mathrm{d}A + \lambda_2 \int_{\delta V} \alpha \, \mathrm{d}A \right\}, \quad (1)$$

where λ_1 and λ_2 denote weighting coefficients, c represents a scalar field of surface consistency and the consistency-scaled vector field $c\mathbf{N}$ contains information about both the local probability of surface presence and the local normal information for the points in the volume and α represents a regularization parameter. The first term in the functional (1) is minimized for high consistency values and a surface which is perpendicular to the observed normals $\mathbf{n}$. The second part represents a regularization term which enforces a minimal surface area to avoid overfitting by increasing the cost for oscillating surfaces. Similar to [35], the global optimization of this functional can be mapped to the optimization of the continuous min-cut functional [40]

$$\min_\lambda \left\{ \int_\Omega (1-\lambda) C_s + \lambda C_t + C \left| \nabla \lambda \right| \mathrm{d}V \right\} \quad (2)$$

via specifying $C = \lambda_2 \alpha$, $C_s = \lambda_1 \max\{0, \mathrm{div}(c\mathbf{N})\}$ and $C_t = \lambda_1 \max\{0, -\mathrm{div}(c\mathbf{N})\}$. We choose this formulation as it provides efficiency concerning memory consumption and alleviates metrification errors.

After describing the utilized setup in the following Section, we describe the technique to acquire and integrate the normal information in Section 5.

4. Acquisition System and Calibration

For the acquisition, we use a turntable-based setup illustrated in Figure 2, where eleven cameras with a resolution of $2{,}048 \times 2{,}048$ pixels are positioned on a vertical arc. The calibration of the cameras and the turntable axis is performed using a rotating three-dimensional calibration target with robustly detectable markers. Similar to *e.g.* [16, 27], we use a monitor-based shape-from-specularity approach to simulate a dense illumination area. Two static displays with resolutions of $2{,}048 \times 1{,}152$ and $2{,}560 \times 1{,}600$ pixels are placed close to the objects for displaying patterns. Gray code patterns and their inverses are used for the unique identification of the reflection of each screen pixel on the mirroring surface with a small number of acquired images. For illuminating the object surface as completely as possible, we place the object onto the display of an Asus TF300T-1E031A tablet with a resolution of $1{,}280 \times 800$ pixels, which is on top of the turntable and also used for displaying patterns. Both the monitor displays and the tablet display need to be placed in a way that provides a good coverage

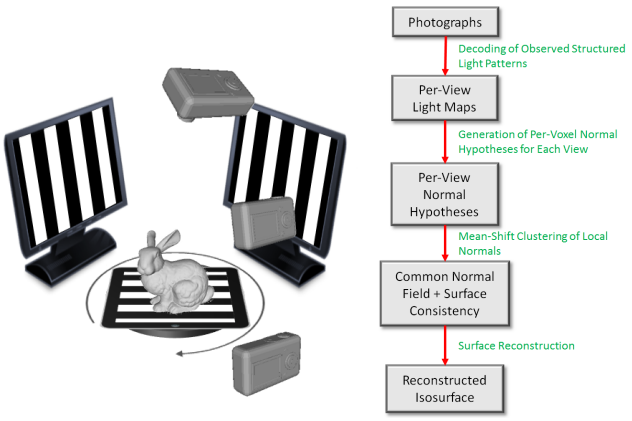

Figure 2: (left) Sketch of the utilized setup: The screens successively display the series of patterns for each rotation of the turntable. The reflected pattern on the object surface is observed by the cameras. For illustration purposes, only three of the eleven employed cameras are drawn. (right) Block diagram of the proposed method.

of the sphere of possible reflection directions. Additionally, we found it important that the tablet is stable enough to support the object weight in case of placing the object on it, *i.e.* tablets with hard glass surfaces are more suitable. In turn, this results in interreflections which have to be taken into account during the reconstruction.

For calibrating the positions of the utilized displays, we use the decoded pattern information observed in the images of the involved cameras and perform an estimation of the display pixel positions $\mathbf{x}_l$ via triangulation so that the resulting point cloud represents (a part of) the display. From the decoded bits for each of the m points in the point cloud, it is possible to uniquely determine its offset $\mathbf{u}_l = [u_l, v_l]^T$ from the origin $\mathbf{o}$ of the display frame which we consider to be at the upper-left. Using this information, we can derive the coordinate frame of the screen consisting of the origin $\mathbf{o}$ and the spanning vectors $\mathbf{a}$ (parallel to the display width) and $\mathbf{b}$ (parallel to the display height) via optimizing $Q = \sum_{l=1}^{m} (\mathbf{x}_l - (\mathbf{o} + u_l \mathbf{a} + v_l \mathbf{b}))^2$. The resulting linear system is solved using least squares minimization. Given the screen calibration, we can directly determine the 3D location of a pixel on the screen by considering its bit sequence.

For the calibration of the screens, it is not necessarily required to see the complete screens in the camera images as several parts of the displays seen in different cameras are sufficient. While our calibration method requires the monitor to be close to the object, this is eventually desirable for the measurement to cover a larger part of the mirror surface with the projected patterns and reduce the influence of light fall-off.

5. Multi-View Shape-from-Coded-Illumination

For bringing classical shape-form-specularity techniques to the multi-view scenario, we first discuss the utilized encoding of the illumination patterns as well as the problems occurring due to surface curvatures which we solve via a fuzzy decoding of the patterns. Subsequently, we describe how the decoded information is used to generate normal hypotheses from which the normal field required in the optimization (1) and the surface consistency are derived. The block diagram of our method is shown in Figure 2.

5.1. Coded Illumination

For encoding the illuminations coming from the displays, we use Gray code patterns which enable a robust decoding. Additionally, similar to the approach presented in [33], we take the inverse patterns for increasing robustness. For decoding the displayed bit sequences, we compare the intensity values observed at each pixel $\mathbf{u}$ in the pair consisting of image $\mathcal{I}_{i,j,k,q}$ seen while displaying pattern $\mathcal{P}_q$ and image $\bar{\mathcal{I}}_{i,j,k,q}$ seen while displaying its inverse pattern $\bar{\mathcal{P}}_q$. If the difference is below a certain threshold, we mark the decoded bit as unreliable. We use $\left|\mathcal{I}_{i,j,k,q} - \bar{\mathcal{I}}_{i,j,k,q}\right| < 0.1\, \mathcal{I}_{i,j,k,0}$, where $\mathcal{I}_{i,j,k,0}$ represents the photo taken under illumination by the fully lit pattern.

As each pixel on the displays can be uniquely encoded and its 3D position on the screen is known from the screen calibration, observed codewords can directly be related to the corresponding 3D positions on the screen. Hence, we generate a light map [3, 9] for each individual camera under each rotation angle k of the turntable and under illumination from each display j. These light maps $\mathcal{L}_{i,j,k}$ assign to each pixel in the camera image the light source position. In general, there will not be observations for all the pixels. The reason for this is that, depending on the shape of the object and the position of the illuminant, only a part of the surface will reflect patterns towards the camera.

Interreflections introduce outliers in the light maps. In addition, depending on the curvature of the mirroring surface and the differing relative distances to the display pixels or other effects, such as non-ideal or spatially varying reflectance properties, it is usually not possible to decode the complete bit sequence correctly. High-frequency patterns might appear blurred on the object surface which has already been observed in *e.g.* [16, 15, 3] and it is not possible to decide if pattern $\mathcal{P}_q$ or its inverse $\bar{\mathcal{P}}_q$ has been displayed. As a consequence, we introduce a fuzzy decoding. The basic idea is to only use the reliably decoded bits per pixel to identify the corresponding display area which illuminated this pixel. If less bits can be reliably decoded, the ambiguity in the region of the display which illuminated the pixel

increases. The corresponding light source position is determined as the center of this reliably decoded region.

To address noisy decodings in the light maps which could represent problems for the calculation of normals and, hence, also for the normal field integration algorithm, we additionally perform a subsequent filtering step. In this step, all decoded labels with less than t_{bits} reliably decoded bits for both horizontal and vertical stripe patterns are discarded. For calibrating the screens, we use $t_{bits} = 9$ as a very accurate decoding is possible. During the reconstruction, we use $t_{bits} = 5$. Furthermore, we also consider for each image pixel per series of patterns the average of the individual contrasts observed for the individual patterns and their inverses to filter out unreliable decodings. In principle, the quality of the decodings can be used as weights for the quality of the normals derived from them. However, in the scope of this paper, we did not investigate this.

5.2. Generation of Normal Hypotheses

The light maps described in the previous subsection are used to derive information about surface normals. As our setup violates the assumtion of distant illumination and the object surface is unknown a priori, the ambiguity concerning the depth of the surface along the view directions for the individual cameras cannot be discarded as in the case of far-field illumination. In our variational formulation, we therefore consider a volumetric representation to resolve this problem. In particular, the normal hypotheses are calculated separately for all the points along the view direction per pixel in each camera similar to [6] by utilizing the information stored in the light maps. For each point $\mathbf{x}$ in the volume and each combination of camera index $i = 1, \ldots, \kappa_c$ screen index $j = 1, \ldots, \kappa_s$ and rotation index $k = 1, \ldots, \kappa_r$, we compute a normal estimate $\mathbf{n}_{i,j,k}(\mathbf{x})$. Assuming that the object remains fixed and cameras and displays are rotated, we consider the coordinate $\mathbf{x}$ relative to the turntable. Therefore, we obtain light directions $\mathbf{l}_{j,k}(\mathbf{x})$ and view directions $\mathbf{v}_{i,k}(\mathbf{x})$ which depend on the position in the volume $\mathbf{x}$ and both on the rotation index k and the screen index j or camera index i respectively. Following the law of reflection, we obtain the normal estimate $\mathbf{n}_{i,j,k}(\mathbf{x})$ as the bisector between $\mathbf{l}_{j,k}(\mathbf{x})$ and $\mathbf{v}_{i,k}(\mathbf{x})$. At points close to the surface, normal hypotheses derived for different camera/screen/rotation configurations, for which the corresponding points are visible, have only a small variance and almost coincide with the true surface normal. In contrast, hypotheses contradict each other at points distant to the true surface.

However, as the cameras might directly observe certain parts of the displays as well, the light maps do not only contain information about the object to be reconstructed. For the reconstruction of the object geometry, these regions in the light maps should not be propagated into the volume in the process of generating normal hypotheses. For this reason, our method also analyzes the 3D distance between the intersection of the view rays with the plane of the active display and the light source position stored in the light map. If this distance is small (we use a threshold of 3mm), it is a hint that the information stored in the light map belongs to the screen geometry and can be masked out.

5.3. Multi-View Normal Field Integration and Surface Consistency Estimation

The result of the normal calculation step is a set of normal fields assigned to the involved capture configurations (i, j, k). These individual fields need to be combined to one common normal field which contains information about the best local normal and the surface consistency.

After combining the information in the volume of interest, we have several normal votes for the different points in this volume. For finding the true surface, we assume that, at a certain location $\mathbf{x}$ on the object surface, the normal hypotheses from the different cameras agree with each other and with the true surface normal. In contrast, normal estimates from the different configurations (i, j, k) will contradict each other further away from the surface. However, due to effects such as outliers, noise, non-ideal calibration or the discretization of the volume, perfectly matching normals will hardly occur in real-world scenarios. Therefore, we can consider the observed normals as samples from an underlying probability distribution. Since the non-occluded normals should agree up to a small variance in the vicinity of the true surface, the underlying distribution should have a global maximum centered around the surface normal. Furthermore, its variance can be regarded as a measure for surface consistency. Similar measures have been used in [6, 27] for reconstructing highly specular and mirroring objects. As the information about visibility of points w.r.t. the involved cameras is unknown, we have to also take into account that several of the normals actually come from an occluded view in additition to the noise and outliers.

Modeling the probability density of normals under occlusions is challenging as it depends on the geometry of the considered object as well as on the placement of the involved cameras and screens. Therefore, we do not model the probability density function (pdf) via a parametric model but instead only make the simplifying assumption that the density is highest for the actual surface normal. This assumption is warranted as the actual surface normal is consistent over all views where the respective surface point has been observed, whereas the outliers should not be consistent over several views. Under this assumption, finding the normal direction corresponds to finding the largest mode of the pdf. For this reason, we decided to use mean-shift clustering [10] as a non-parametric technique as it neither requires assuming a model nor creates discretization artifacts. We

therefore define the pdf as

$$p_{\mathbf{x}}(\mathbf{n}) = \frac{1}{\kappa_c \kappa_s \kappa_r h^3} \sum_{i,j,k} K\left(\frac{\|\mathbf{n} - \mathbf{n}_{i,j,k}(\mathbf{x})\|}{h}\right), \quad (3)$$

and set the local normal estimate $\mathbf{N}(\mathbf{x}) = \arg\max_{\mathbf{n}} p_{\mathbf{x}}(\mathbf{n})$ to the centroid of the highest mode of the pdf. Furthermore, we use the density at the centroid as a surface consistency measure which we denote with $c(\mathbf{x}) = p_{\mathbf{x}}(\mathbf{N}(\mathbf{x}))$.

In eq. (3), K represents the kernel function with bandwidth h. We experimented with both the Epanechnikov kernel and the Gaussian kernel and found the latter to result in a more accurate reconstruction. We heuristically determined $h = 0.03$ which worked for all our datasets, but did not perform a complete evaluation on the sensitivity to this parameter. As an alternative, it is also possible to consider normal histograms. Then, the highest mode of the pdf corresponds to the bin with the maximum count. Though this would be faster, in our experiments, we did not reach the quality of the reconstructions when using mean-shift clustering.

5.4. Surface Reconstruction

After calculating the estimates for the common volumetric normal field and the surface consistency as described before, we adapt the iterative optimization procedure presented in [35] to our setting. After an initialization of the utilized octree at a coarse level, the grid is successively refined according to the local surface consistency estimates in the volume. In a subsequent iterative process, the memory efficient continuous min-cut [40] is applied for a global optimization per iteration. In a final step, the resulting binary indicator function is smoothed inspired by the technique presented in [7].

6. Experimental Results

We evaluate our technique in two steps. To demonstrate the robustness of our reconstruction framework, we first consider the classical multi-view normal field integration case. Here, we use per-camera normal images as input. In the next step, we show results on mirroring objects. For the experiments mentioned in the paper, we use 264 views ($\kappa_c = 11$ cameras are mounted on an arc and the turntable is rotated in steps of $15°$, *i.e.* $\kappa_r = \frac{360°}{15°} = 24$).

In a first experiment, we consider 3D reconstruction from several per-view normal fields. For this purpose, we have acquired a painted mask made of clay and estimate for each view an independent normal map using classical single-view photometric stereo [36]. Subsequently, the integration is performed using our variational formulation. We use the assumption of far-field illumination but with our technique it would also be possible to relax this assumption by computing an individual normal at each point in the volume. As the assumption of Lambertian surface reflectance

(a) Camera image. (b) Reconstruction.

Figure 3: Results on a photometric stereo dataset: In particular, the painted regions of the clay mask exhibit specularities which leads to a violation of the assumption regarding Lambertian reflectance behaviour. Nevertheless, the reconstruction preserves the shape in these regions.

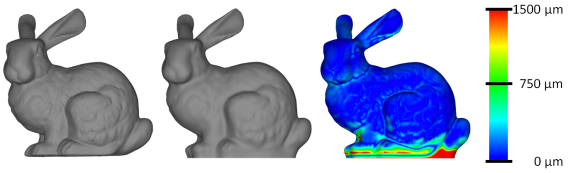

Figure 4: Stanford Bunny model, reconstructed model and visualization of the reconstruction error. Except for the bottom region, where almost no observations have been available, the reconstruction on an adaptive grid of level nine (at which the voxel edge length is approx. 250μm) fits to the original model.

is violated due to the presence of effects such as specular reflection, shadows and inter-reflections on the mask surface, normal estimation based on linear least-squares fitting is prone to errors. Therefore, we use a simple outlier rejection to remove the influence of too bright or too dark regions in the least-squares fitting. The reconstructed model is shown in Figure 3. Applying a more sophisticated photometric stereo technique would probably improve the reconstruction quality. We also show results for a synthetic test case in the supplementary material where normal fields are directly generated from the object geometry using a normal shader in OpenGL. The results demonstrate that fine surface details are well-preserved in the reconstruction.

For mirroring objects, we first consider a synthetic test case where we represent each display via a plane textured according to the patterns of the Gray code sequence. The scene is rendered using conventional ray tracing using 64 samples per pixel to accurately simulate the blurring in curved regions due to limited camera resolution. We use a camera resolution of $2{,}048 \times 2{,}048$ pixels and simulate $\kappa_s = 2$ screens which results in $\kappa_s \cdot \kappa_c \cdot \kappa_r = 528$ light maps. Figure 4 shows a comparison of the original model and the reconstruction. For evaluating the robustness of our

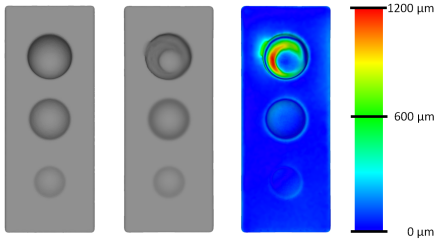

Figure 5: Block model with pits of increasing depth, reconstructed model and visualization of the reconstruction error (level nine reconstruction, *i.e.* the voxel edge length is approx. 250μm at level nine).

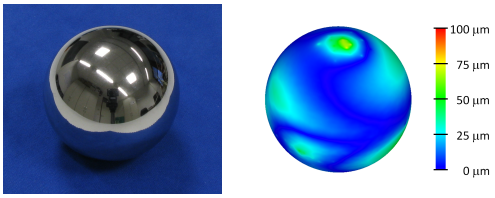

Figure 6: Reconstruction of a mirroring sphere (object and visualization of the Hausdorff distance between the reconstructed model and a sphere according to the sphere specifications).

approach w.r.t. interreflections, we also considered a synthetic, mirroring block with pits of increasing depth, where the proportion of multiple-bounce observations gradually increases. The reconstruction results are shown in Figure 5. Finally, we evaluate our technique on two mirroring, real-world objects. For obtaining information about the accuracy of our approach, we have measured a precisely manufactured sphere with a radius of 25 mm (using $\kappa_s = 2$ screens) and compare the reconstructed model to an ideal sphere of fixed radius whose center is determined via a least squares fit. The error is measured via the Hausdorff distance and shown in Figure 6. The root mean square error of the reconstruction is 20μm which is considerably lower than the edge length of a voxel (approx. 200μm) on the utilized maximum octree level. In comparison, one image pixel corresponds to approx. 150μm at the distance of the object.

Furthermore, we test the robustness of our approach on objects with a more complex surface geometry such as self-occluded parts and concavities which lead to interreflections. For this purpose, we have acquired a mirroring bunny figurine. The reconstruction in Figure 1 clearly indicates the possible reconstruction accuracy. Using additional octree levels could further improve the reconstruction at the cost of higher computation effort and memory requirements.

During our experiments, we mainly start with an initial subdivision on level seven and perform each times three surface adaptions before going to the next higher octree level. On a Intel Xeon E5654 CPU with 2.4 GHz, our level nine reconstruction for $\kappa_r = 24$, $\kappa_c = 11$ and $\kappa_s = 3$, as used for the real-world bunny figurine, which leads to 792 observed individual normal fields, requires approximately 12 hours, while the acquisition took approx. 2 hours.

The results shown in this section indicate the potential of normal-based surface reconstruction. In contrast to the previously presented multi-view normal field integration approaches in [8, 12], our method is robust enough to deal with real-world data in the presence of noise and outliers. However, regions such as concavities with a certain orientation to the displays, under which no information can be observed, cannot be accurately reconstructed.

7. Conclusions

In this paper, we have presented a novel, robust multi-view normal field integration technique for reconstructing the full 3D shape of mirroring objects. Based on coded illumination, our technique derives several normal hypotheses for each point of the considered volume. From these hypotheses, both the most likely local surface normal and a local surface consistency estimate are computed. In our experiments, we have demonstrated that our method yields accurate 3D reconstructions of highly-specular objects even in the presence of occlusions.

Current limitations can be found when considering deep concavities or other parts of the surface, where no information has been observed. Resolving these problems is challenging as it would require considering multiple scattering. Since the underlying optimization technique is independent of the source of the estimated normals, we would like to extend our method to objects which are only partially mirroring and also exhibit other surface reflectance behavior.

Acknowledgements: The research leading to these results was funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 323567 (Harvest4D); 2013-2016.

References

[1] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Towards a theory of shape from specular flow. In *Proc. ICCV*, pages 1–8, 2007.

[2] J. Balzer, S. Holer, and J. Beyerer. Multiview specular stereo reconstruction of large mirror surfaces. In *Proc. CVPR*, pages 2537–2544, 2011.

[3] J. Balzer and S. Werling. Principles of Shape from Specular Reflection. *Measurement*, 43:1305–1317, 2010.

[4] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *IJCV*, 72:239–257.

[5] M. Beljan, J. Ackermann, and M. Goesele. Consensus multi-view photometric stereo. In *DAGM/OAGM Symposium*, pages 287–296, 2012.

[6] T. Bonfort and P. Sturm. Voxel carving for specular surfaces. In *ICCV*, pages 691–696, 2003.

[7] F. Calakli and G. Taubin. Ssd: Smooth signed distance surface reconstruction. *Comput. Graph. Forum*, 30(7):1993–2002, 2011.

[8] J. Y. Chang, K. M. Lee, and S. U. Lee. Multiview normal field integration using level set methods. In *CVPR*, pages 1–8, 2007.

[9] T. Chen, M. Goesele, and H.-P. Seidel. Mesostructure from specularity. In *CVPR*, volume 2, pages 1825 – 1832, 2006.

[10] Y. Cheng. Mean shift, mode seeking, and clustering. *TPAMI*, 17(8):790–799, 1995.

[11] E. N. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing*, 18(4):309 – 328, 1982.

[12] Z. Dai. A markov random field approach for multi-view normal integration. Master-Thesis, University of Hong Kong, 2009.

[13] A. Delaunoy, E. Prados, and P. N. Belhumeur. Towards full 3D Helmholtz stereovision algorithms. In *ACCV 2010 - Volume Part I*, pages 39–52, 2011.

[14] C. H. Esteban, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *TPAMI*, 30(3):548–554, 2008.

[15] Y. Francken, T. Cuypers, T. Mertens, and P. Bekaert. Gloss and normal map acquisition of mesostructures using gray codes. *Advances in Vis. Computing*, pages 788–798, 2009.

[16] Y. Francken, T. Cuypers, T. Mertens, J. Gielis, and P. Bekaert. High quality mesostructure acquisition using specularities. *CVPR*, pages 1–7, 2008.

[17] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *ICCV*, volume 1, pages 341–348, 2005.

[18] A. Hertzmann and S. M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *TPAMI*, 27(8):1254–1264, 2005.

[19] T. Higo, Y. Matsushita, and K. Ikeuchi. Consensus photometric stereo. In *CVPR*, pages 1157–1164, 2010.

[20] B. K. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, Cambridge, MA, USA, 1970.

[21] I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich. Transparent and specular object reconstruction. *Computer Graphics Forum*, 29(8):2400–2426, 2010.

[22] K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *TPAMI*, 3(6):661–669, 1981.

[23] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. *IJCV*, 76(1):13–29, 2008.

[24] M. Liu, K.-Y. K. Wong, Z. Dai, and Z. Chen. Specular surface recovery from reflections of a planar pattern undergoing an unknown pure translation. In *ACCV (2)*, pages 137–147, 2010.

[25] A. Maki and R. Cipolla. Obtaining the shape of a moving object with a specular surface. In *Proc. BMVC*, pages 39.1–39.10, 2009.

[26] S. Nayar, L. Weiss, D. Simon, and A. Sanderson. Specular Surface Inspection using Structured Highlight and Gaussian Images. *Transactions on Robotics and Automation*, 6(2):208–218, Apr 1990.

[27] D. Nehab, T. Weyrich, and S. Rusinkiewicz. Dense 3D reconstruction from specularity consistency. In *CVPR*, 2008.

[28] S. Roth and M. J. Black. Specular flow and the recovery of surface structure. In *Proc. CVPR*, volume 2, pages 1869–1876, 2006.

[29] A. C. Sanderson, L. E. Weiss, and S. K. Nayar. Structured highlight inspection of specular surfaces. *TPAMI*, 10(1):44–55, 1988.

[30] A. C. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. K. Agrawal. Specular surface reconstruction from sparse reflection correspondences. In *CVPR*, pages 1245–1252, 2010.

[31] S. Savarese, M. Chen, and P. Perona. Local shape from mirror reflections. *IJCV*, 64(1):31–67, 2005.

[32] M. Tarini, H. P. A. Lensch, M. Goesele, and H.-P. Seidel. 3d acquisition of mirroring objects using striped patterns. *Graph. Models*, 67(4):233–259, 2005.

[33] M. Trobina. Error model of a coded-light range sensor. Technical report, Communication Technology Laboratory, ETH Zentrum, Zurich, 1995.

[34] Z. Wang and S. Inokuchi. Determining shape of specular surfaces. In *Scandinavian Conference on Image Analysis*, pages 1187–1194, 1993.

[35] M. Weinmann, R. Ruiters, A. Osep, C. Schwartz, and R. Klein. Fusing structured light consistency and helmholtz normals for 3d reconstruction. *BMVC*, Sept. 2012.

[36] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

[37] C. Wu, Y. Liu, Q. Dai, and B. Wilburn. Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. *Transactions on Visualization and Computer Graphics*, 17(8):1082–1095, 2011.

[38] M. Yamazaki, S. Iwata, and G. Xu. Dense 3d reconstruction of specular and transparent objects using stereo cameras and phase-shift method. In *ACCV (2)*, pages 570–579, 2007.

[39] S. K. Yeung, T.-P. Wu, C.-K. Tang, T. F. Chan, and S. Osher. Adequate reconstruction of transparent objects on a shoestring budget. In *CVPR*, pages 2513–2520, 2011.

[40] J. Yuan, E. Bae, and X.-C. Tai. A study on continuous max-flow and min-cut approaches. In *CVPR*, 2010.

[41] J. Y. Zheng and A. Murata. Acquiring a complete 3d model from specular motion under the illumination of circular-shaped light sources. *TPAMI*, 22(8):913–920, 2000.

[42] T. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. In *IJCV*, pages 869–884, 2002.

[43] A. Zisserman, P. Giblin, and A. Blake. The information available to a moving observer from specularities. *Image and Vision Computing*, 7(1):38–42, 1989.