

# Fusing Structured Light Consistency and Helmholtz Normals for 3D Reconstruction

Michael Weinmann  
mw@cs.uni-bonn.de

Roland Ruiters  
ruiters@cs.uni-bonn.de

Aljosa Osep  
osep@cs.uni-bonn.de

Christopher Schwartz  
schwartz@cs.uni-bonn.de

Reinhard Klein  
rk@cs.uni-bonn.de

Institute of Computer Science II  
University of Bonn  
Germany

---

## Abstract

In this paper, we propose a 3D reconstruction approach which combines a structured light based consistency measure with dense normal information obtained by exploiting the Helmholtz reciprocity principle. This combination compensates for the individual limitations of techniques providing normal information, which are mainly affected by low-frequency drift, and those providing positional information, which are often not well-suited to recover fine details. To obtain Helmholtz reciprocal samples, we employ a turntable-based setup. Due to the reciprocity, the structured light directly provides the occlusion information needed during the normal estimation for both the cameras and light sources. We perform the reconstruction by solving one global variational problem which integrates all available measurements simultaneously, over all cameras, light source positions and turntable rotations. For this, we employ an octree-based continuous min-cut framework in order to alleviate metrification errors while maintaining memory efficiency. We evaluate the performance of our algorithm both on synthetic and real-world data.

## 1 Introduction

The faithful reconstruction of 3D shape is one of the classic goals in the field of computer vision. Important applications are quality assurance, reverse engineering, digital preservation or computer graphics. For these purposes, highly detailed and hole-free 3D geometries are desirable. This task is especially challenging for objects made of materials with complex non-Lambertian reflectance behaviour.

A common application is the photo-realistic visual reproduction of real world objects. Here, it is additionally necessary to acquire the reflectance characteristics. For an automated capture of such datasets, setups capable of taking images of the object under varying view- and illumination conditions have been developed. To capture the corresponding 3D geometry, laser scanners or structured light are employed. However, such 3D reconstructions from

triangulation either suffer from noise or over-smoothing. In contrast, reconstruction based on normal integration would be capable of preserving high-frequency surface details but is prone to low-frequency drift due to the accumulation of errors. Combinations of these two techniques have been proposed to overcome the respective problems. In this paper, we will explore the combination of structured light with normal information estimated by utilizing the Helmholtz reciprocity principle. Both techniques make only modest assumptions about surface reflectance. As a result, they can be used to capture nearly all opaque objects; the only notable exception being objects with extreme specularities, such as mirrors. Additionally, this combination is especially useful in the aforementioned setting of photo-realistic reproduction, since the images needed for reflectance reconstruction can be used for the Helmholtz normal estimation. The structured light directly provides information about occlusion and shadowing that can be utilized in the Helmholtz stereopsis.

Most existing reciprocal setups, i.e. devices for which the view- and light-direction can exactly be exchanged, rely on moveable cameras and light sources. We employ instead a device with fixed lights, cameras and projectors, using a turntable as the only moving part. This simplifies calibration and hardware complexity considerably. By using a symmetric layout of the light sources, such a setup can be utilized to obtain pairs of Helmholtz reciprocal samples (see Figure 1).

Since the turntable rotates during the acquisition, the projected patterns move over the surface of the object. Therefore, direct triangulation of codewords can only be performed per rotation. Instead of using all cameras that have seen one location on the surface to compute a single consistent point, an independent triangulation is performed for each rotation, resulting in multiple potentially contradicting solutions. Regaining the fine surface details from the resulting noisy point cloud, however, is a non-trivial task.

To overcome this limitation and obtain one globally consistent reconstruction integrating the information over all rotations, we use a variational approach which combines a consistency term derived from structured light with the Helmholtz normals. We solve the respective optimisation problem via an octree-based continuous min-cut framework which is memory efficient and alleviates metrification errors. To compensate for the discretisation artefacts from the min-cut, a smooth signed distance function is then computed from the resulting binary labelling, again taking the reconstructed normals into account. Finally, the reconstruction result is derived from this smooth signed distance.

In summary, the key contributions of our work are i) a variational reconstruction technique combining structured light and normals from Helmholtz reciprocity, ii) a structured light based consistency measure which allows to combine several structured light measurements although the object was moved with respect to the projector, iii) a memory-efficient octree based continuous min-cut solver, iv) a final refinement step based on a smoothed signed distance function derived from the min-cut and the Helmholtz normals and v) a turntable-based symmetrical setup for acquiring reciprocal image pairs.

## 2 Previous Work

A variety of different principles has been developed for acquiring 3D shapes, for which surveys can be found in [25] and [10]. Although passive methods such as multi-view stereo [27] exist, active techniques that directly influence the lighting in a scene are capable of acquiring higher quality reconstructions and deal with a significantly larger range of surface reflectance characteristics. Among these, structured light based methods reach an impres-

sive reconstruction accuracy [9] and are mainly limited by the resolution of the projectors. Several approaches try to alleviate this restriction, e.g. by combining codeword information within a multi-camera and multi-projector setting [30] or by using lens-shifted structured light [24]. However, a major drawback of all methods relying on the triangulation of feature correspondences is that this triangulation entails high-frequency noise in the final reconstruction. This noise is mainly due to inaccurate calibration or feature localization. As a result, fine details might not be captured.

In order to overcome limitations of triangulation-based techniques, several investigations focus on fusing consistency or triangulated positional information with complementary visual cues, e.g. normal orientation. Many approaches employ deformable models to guide an iterative surface evolution by optimizing a cost function to enforce consistency. For this, the surface is often represented in a volumetric way [9], [15] or by polygonal meshes [23], [9], [10], [22], [0], [8], [0], [52]. These methods rely on a good initialization and evolve the surface via a variational problem using gradient flows. However, such surface evolution techniques can get stuck in local minima. In addition, mesh-based techniques are not capable of handling topological changes of the surface geometry such as self-intersecting regions in the mesh. This can be overcome by hybrid approaches [53].

In contrast, global optimization strategies such as graph cuts [28], [14], [54], [30], [20], [12] or convex optimization [18] overcome the local minima problem by considering a cost function defined on the whole volume for surface extraction. Due to their discrete nature, graph cut approaches necessitate the use of high-resolution volumetric grids for obtaining high-quality reconstructions. In addition, a high grid-connectivity is desirable for reducing metrification errors. Incorporating these aspects comes at the expense of a drastically increasing memory consumption. This problem can be alleviated to some extent by employing adaptive grid structures such as octrees. Convex optimization, in contrast, does not suffer from metrification errors. Thus, shifting variational problems from the discrete to the continuous domain has recently gained attention [6], [19], [35]. Although these techniques are more memory efficient, storing the data on a regular grid still imposes severe resolution restrictions. In this paper, we adapt the continuous min-cut proposed in [35] to an octree structure.

For a reliable reconstruction of fine surface details, the accuracy of the normals is a crucial factor. However, obtaining accurate normals for the wider range of non-Lambertian opaque objects is a challenging task. Therefore, several techniques rely on simplifying assumptions such as either Lambertian or purely specular surface reflectance, which limits the range of objects that can be acquired. In contrast, obtaining accurate normals for opaque objects has been investigated by exploiting the duality between cameras and light sources via the Helmholtz reciprocity [56], which requires an accurate co-location of the cameras and light sources. The only assumption is that the surface materials present in the scene have a reciprocal bi-directional reflectance distribution function (BRDF). Recently, this principle has also been used in a multi-view setting in [8].

### 3 Setup

Our measurement device is designed to enable the acquisition of images under structured light and varying view- and illumination-conditions, explicitly allowing the capture of reciprocal image pairs. Setups with similar functionality have been addressed in several publications (e.g. [13], [26]) for the capture of 3D surfaces with reflectance information. In these

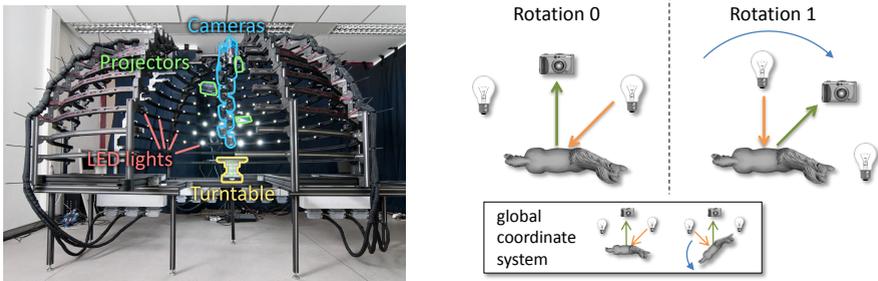


Figure 1: **The reciprocal measurement setup** Left: one quarter of the light-dome is slid open to provide better view on the inside. Right: Illustration of symmetrical light setup. The lower images illustrate the actual rotation of the object on the turntable in the measurement setup. The upper images are shown in the object coordinate system and demonstrate how this rotation leads to reciprocal image pairs.

works, however, the Helmholtz principle was not utilized.

In our setup depicted in Figure 1, 11 cameras ( $2048 \times 2048$  pixel) are mounted on a vertical arc and oriented towards a turntable. For acquiring positional information of the object shape, 4 projectors with a resolution of  $800 \times 600$  pixels are installed in the vicinity of the cameras. During the measurement process, these are used to project Gray code patterns onto the object surface. 198 LED lights are additionally distributed over a hemispherical gantry. Care has been taken that these LEDs are placed in a symmetrical manner w.r.t. the turntable axis and the cameras, so that we can obtain reciprocal image-pairs by rotating the turntable (as illustrated in Figure 1).

The geometric calibration of cameras and projectors (pinhole camera model including distortion) is obtained via a structured light based bundle adjustment [64]. For estimating the turntable axis, a reference calibration target with easy-to-detect features is placed onto the turntable and the features are tracked. In addition, light source positions are estimated using four mirroring spheres which can be rotated with the turntable. For this, highlights on the spheres are detected under illumination by the individual LEDs and the rays which are reflected from the spheres are triangulated. The setup is also fully radiometrically calibrated via standard reflectors for the cameras and LED light sources so as to be able to exploit the Helmholtz reciprocity. This necessitates that the image-pair measurements are conducted in high dynamic range (HDR). In addition, we also use HDR imaging for the structured light to facilitate robust pattern-decoding (see [64]).

With this setup, we acquire a number of images of an object. For each view- and rotation-configuration, the data consists of radiometrically corrected colour values under several illuminations and the per-pixel codewords decoded from the structured light.

## 4 Variational Formulation

We formulate the reconstruction of the object surface  $\delta V$  as a variational problem. The formulation depends on a vector field of normals  $\mathbf{H}$  and three scalar fields defined on the continuous volume  $\mathbb{R}^3$ : the consistency measure  $c$ , the outside count  $o$  and the visibility count  $v$ . We use  $c(\mathbf{x})$  to denote the number of camera-projector consistent codewords detected at  $\mathbf{x} \in \mathbb{R}^3$ ,  $o(\mathbf{x})$  for the number of cases in which structured light triangulation has determined

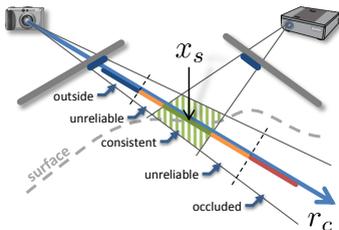


Figure 2: Structured light consistency terminology illustration

The cones through the decoded codeword in the camera-image and the known source-pixel in the projector plane intersect to form a consistency volume. The object surface intersects this volume. All positions  $\mathbf{x}$  along the viewing ray  $r_c$  are classified as being in front of or behind the surface.

$\mathbf{x}$  to be outside of the object and  $v(\mathbf{x})$  for the total number of configurations in which  $\mathbf{x}$  was visible from the camera. From these, we derive visibility-normalized versions  $\hat{c} = \frac{c}{v}$  and  $\hat{o} = \frac{o}{v}$ . The resulting consistency-weighted vector field  $c\mathbf{H}$  has large magnitude in the vicinity of the surface, is aligned perpendicular to the surface and diminishes further away. To find the object interior  $V \subset \mathbb{R}^3$ , we seek to solve the following problem

$$\min_V E(V) = -\lambda_1 \underbrace{\int_{\delta V} \langle c\mathbf{H}, \mathbf{n} \rangle dA}_{E_1} + \lambda_2 \underbrace{\int_V \hat{o} dV}_{E_2} + \lambda_3 \underbrace{\int_{\delta V} (\alpha - \hat{c}) dA}_{E_3}, \quad (1)$$

where  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are relative weights of the individual terms and  $\alpha > 1$  denotes a constant determining the the minimum regularization strength within a completely consistent region. The first term  $E_1$  considers the flux of the vector field  $c\mathbf{H}$  through the object surface. This term is minimized by a surface running perpendicular to the reconstructed Helmholtz normals  $\mathbf{H}$  and in regions with a high consistency  $c$ . The second term  $E_2$  is used as an outside constraint to penalize regions of large values  $\hat{o}$ . This prevents the algorithm from short-cutting through concavities. The last term  $E_3$  represents a regularization term and enforces a minimal surface. This penalty is weighted with the consistency  $\hat{c}$  obtained from the structured light.

**Structured Light Consistency** For each point  $\mathbf{x} \in \mathbb{R}^3$  and all combinations  $(i, j, k)$  of rotations, cameras and projectors, we perform an independent classification. We determine each time whether the point is consistent, lies before or behind the surface seen in the respective pixel in the camera or cannot be reliably classified at all (see Figure 2). We consider a point consistent if the codeword decoded in the camera image  $j$  agrees with the codeword cast by the projector  $k$ . This can easily be checked by projecting the point into both the camera and the projector and comparing the codewords. All points meeting this criterion form the consistency volume depicted in green hatching in Figure 2. When the point does not lie within this region, we classify it according to its position along the ray  $r_c$  from the camera. It can either lie in front of the intersection of  $r_c$  with the surface  $\mathbf{x}_s$ , and thus outside of the object, or behind the intersection, in which case this sample does not provide any information at all. In practice, we distinguish between these cases by triangulating the position  $\mathbf{x}_s$ . This is possible, since we can determine the projector ray  $r_p$  from the codeword which lies at the projection of  $\mathbf{x}$  into the camera image. In order to incorporate the inaccuracy of triangulation, we consider points that are not consistent but still lie closer to  $\mathbf{x}_s$  than a user-selected threshold  $\delta$  as unreliable. In case the triangulation fails completely, since  $r_c$  and  $r_p$  do not intersect or no codeword was found in the camera image, the sample combination is discarded for  $\mathbf{x}$  altogether.

After performing this classification for every combination  $(i, j, k)$ , we count the number

of consistent combinations  $c(\mathbf{x})$ , combinations in which the point was determined to lie in front of the surface  $o(\mathbf{x})$ , and the number of unreliable combinations  $u(\mathbf{x})$ . We then set the visibility to the number of samples that were not discarded ( $v = c + o + u$ ) and derive the normalized consistency and outside probabilities  $\hat{c}$  and  $\hat{o}$ .

**Estimating Normals** At each point, the corresponding normal information is estimated via the Helmholtz principle [66]. This is based on the reciprocity of the BRDF, i.e.  $\rho(\mathbf{v}, \mathbf{l}) = \rho(\mathbf{l}, \mathbf{v})$ , and can be utilized wherever several image pairs are available in which the position of the light source and the camera have been exactly exchanged. Instead of actually moving the light-source and camera, these image pairs can also be acquired by rotating the object and using two light sources, mounted symmetrically around the camera (see Figure 1). Unfortunately, accurately achieving this symmetry in a setup is difficult for practical reasons. For example, the emitter of the light source and the projective centre of the camera are difficult to align. In a larger setup, there are also physical restrictions for the available mounting positions due to the size of the components and the used gantry configuration. Together, these restrictions might result in errors of a few degrees.

We therefore relax the assumption of perfect Helmholtz stereopsis and instead assume that the BRDF is locally smooth enough to allow for barycentric interpolation between three light samples. For this, we first compute a spherical Delaunay triangulation of the set of available light source directions. Given an (idealized) pair of reciprocal camera-/light-positions  $\mathbf{y}$  and  $\mathbf{y}'$ , we obtain the two corresponding directions for each point  $\mathbf{x}$  as  $\mathbf{d}$  and  $\mathbf{d}'$  and find the respective spherical triangles. These provide us with barycentric interpolation weights  $\alpha_i$ ,  $\alpha'_i$  and light directions  $\mathbf{l}_i, \mathbf{l}'_i$  with  $i \in \{1, 2, 3\}$  in such a way that  $\rho(\mathbf{d}, \mathbf{d}') \approx \sum_{i=1}^3 \alpha'_i \rho(\mathbf{d}, \mathbf{l}'_i)$  and  $\rho(\mathbf{d}', \mathbf{d}) \approx \sum_{i=1}^3 \alpha_i \rho(\mathbf{d}', \mathbf{l}_i)$ .

For each of these BRDF samples, we now have actual irradiance measurements  $I_i$  and  $I'_i$  available. We assume here, that these measurements have already been radiometrically corrected to compensate for the light falloff at the point  $\mathbf{x}$  by using the calibration of the measurement setup. Hence, we have  $\rho(\mathbf{d}, \mathbf{l}'_i) = \frac{I_i}{\langle \mathbf{n}, \mathbf{l}'_i \rangle}$  and  $\rho(\mathbf{d}', \mathbf{l}_i) = \frac{I'_i}{\langle \mathbf{n}, \mathbf{l}_i \rangle}$ . Since we assume from the Helmholtz reciprocity that  $\rho(\mathbf{d}, \mathbf{d}') = \rho(\mathbf{d}', \mathbf{d})$ , we estimate the normal by solving the following optimization problem:

$$\min_{\mathbf{n}} E_n(\mathbf{n}) = \frac{1}{\sum_{j=1}^N w^{(j)}(\mathbf{n})} \sum_{j=1}^N w^{(j)}(\mathbf{n}) \left\| \sum_{i=1}^3 \alpha'_i{}^{(j)} \frac{I_i^{(j)}}{\langle \mathbf{n}, \mathbf{l}'_i{}^{(j)} \rangle} - \sum_{i=1}^3 \alpha_i{}^{(j)} \frac{I'_i{}^{(j)}}{\langle \mathbf{n}, \mathbf{l}_i{}^{(j)} \rangle} \right\|^2. \quad (2)$$

The weight terms  $w^{(j)}(\mathbf{n}) = (\langle \mathbf{n}, \mathbf{d}^{(j)} \rangle \langle \mathbf{n}, \mathbf{d}'^{(j)} \rangle)^2$  are included to reduce the relative importance of samples at grazing angles as these measurements are less accurate and the division through the small cosine term would further increase these errors. If either of the two scalar products becomes smaller than a threshold we set  $w^{(j)}(\mathbf{n}) = 0$ . Only samples which were classified as consistent by the structured light computation are included in the computation to avoid the influence of occlusion and shadowing. We solve the resulting non-linear optimization problem with a Levenberg-Marquardt optimizer using the SVD approach proposed in [66] as initialization.

## 5 Implementation

Storing all values needed for the reconstruction in a regular grid would be prohibitively expensive with regard to memory consumption and computational demands, preventing re-

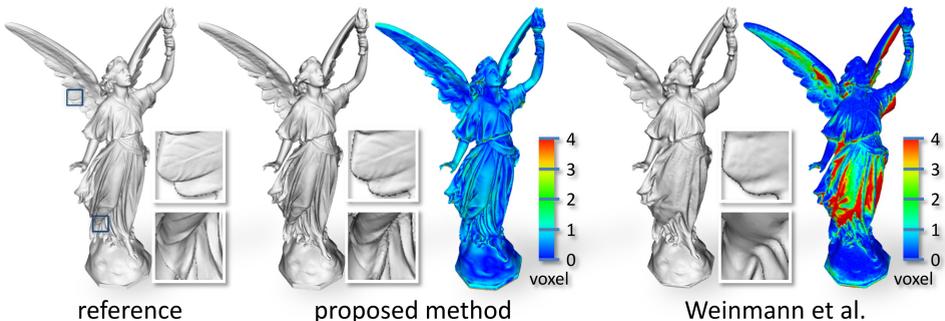


Figure 3: **Comparison on synthetically generated input data.** Our technique preserves fine details, *e.g.* on the wings or the garment, and also reconstructs the concave regions, which are otherwise lost during the Poisson reconstruction of [53]. The plots show the Hausdorff distance to the reference mesh, normalized to the size of an octree cell. Note, our reconstruction remains nearly everywhere within an error of about one octree cell.

constructions at high levels of detail. Therefore, we employ an octree-based data-structure, which adapts the resolution to the shape of the object. For this, we use an iterative algorithm which successively refines the octree based on the previous reconstruction.

**Structured Light Based Octree Initialisation:** Starting with a tree at an initial coarse resolution, we successively refine cells, as long as a common structured light codeword appears in a sufficiently large number of the re-projected footprints of the cell. When the surface runs through the cell, the same codeword should lie within the footprints in a large number of the images. In contrast, the number of coincidentally identical codewords decreases with the size of the cell. The refinement continues up to a maximum initial octree depth.

**Octree Update:** The consistency values  $\hat{\delta}$  and  $\hat{c}$  and the Helmholtz-normals  $\mathbf{H}$  are all evaluated and stored at the corners of each octree cell. Computation of these values is consequently only necessary for the corners newly added by the subdivision process. All previously existing corners in the tree remain unchanged and hence do not require updating.

**Continuous Min-cut:** Having the consistency values and the normals in the volume of interest, we seek to find a globally optimal min-cut, partitioning the volume into inside and outside. Using the Gauss-Ostrogradsky theorem,  $\int_{\partial V} \langle c\mathbf{H}, \mathbf{n} \rangle dA = \int_V \text{div}(c\mathbf{H}) dV$ , we can represent the term  $E_1$  in the functional (1) as an integral over the volume. The total energy can then be directly mapped onto the continuous min-cut functional

$$\min_{\lambda} D(\lambda) = \int_{\Omega} (1 - \lambda)C_s + \lambda C_t + C|\nabla\lambda| dx \quad (3)$$

given in [53]. For this, we set  $C = \hat{c}$ ,  $C_s = \max(0, \text{div}(c\mathbf{H}) - \hat{\delta})$  and  $C_t = \max(0, -\text{div}(c\mathbf{H}) + \hat{\delta})$ . To obtain the desired labeling  $\lambda$  of the volume, we are able to apply the continuous max-flow algorithm proposed in [53] due to the duality between max-flow and min-cut.

In contrast to [53], in which a dense 3D grid was employed, we perform the max-flow computation directly on the proposed octree structure. For this, we compute for each of the cells one value  $C$ ,  $C_s$  and  $C_t$ . To obtain the divergence within the cell, we compute samples of  $c\mathbf{H}$  for each of the cell’s facets by averaging the corresponding values stored at the surrounding corners. Weights based on the Helmholtz normal error  $E_n$  (see Equation

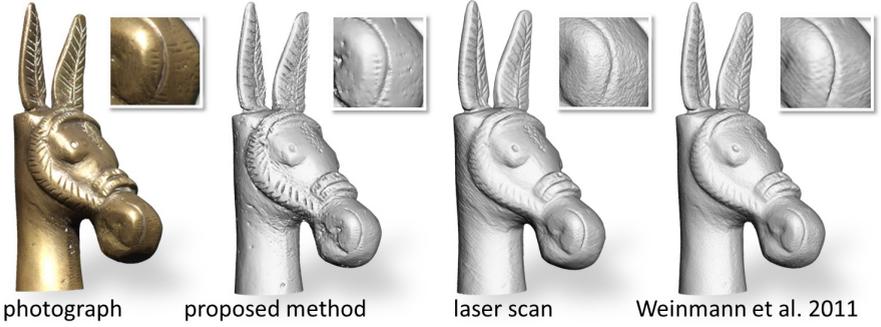


Figure 4: **Comparison of results on a glossy brass figurine.** The laser-scan was created with a high-precision line laser-scanner mounted on a measuring arm with a total accuracy of about  $60\mu\text{m}$ . Our reconstruction shows considerably more fine surface details.

(2)) are employed to decrease the influence of unreliably estimated normals. These samples of the vector field are then used to approximate the flow through each of the facets and, consequently, the average divergence within the cell via the divergence theorem. We also weight these divergences with the average reliability of the Helmholtz normal estimation to avoid large divergences in regions with unreliable normals.

Once the surface has been computed, all octree cells with a distance to the reconstructed surface smaller than twice the cell’s edge length are refined to adaptively increase the octree resolution near to the surface. Furthermore, the maximum allowed octree-level is increased in each iteration to successively refine the reconstruction.

**Smooth Surface Reconstruction:** Especially in regions where the divergence term dominates the min-cut reconstruction, the iso-surfaces of the resulting continuous labelling are not smooth. To infer a smooth final surface, we compute a signed distance function  $f$ . We constrain this function to lie within a band of one octree cell around the reconstructed iso-surface. To incorporate the normal information obtained via Helmholtz stereopsis, we search for a surface corresponding to the vector field  $\mathbf{H}$  that is as smooth as possible. This can be achieved by using an energy function similar to the one proposed in [9]. We utilize the same normal consistency and smoothness terms. However, since we do not have a point cloud as input, we remove the data-term, which penalizes deviations from the points. Instead, we use hard constraints which enforce the implicit function  $f$  to be negative inside of the object and positive on the outside and perform the optimization only within the narrow band. This results in the following optimisation problem:

$$\begin{aligned}
 \min_f \quad & \mu_0 \sum_{i=1}^N w_i \|\nabla f(x_i) - \mathbf{n}_i\|^2 + \mu_1 \frac{1}{|V|} \int_V \|Hf(x)\|^2 dx, \\
 \text{s. t.} \quad & f(x) < -\varepsilon \quad \text{for all } x \text{ at the inner border of the band} \\
 & f(x) > \varepsilon \quad \text{for all } x \text{ at the outer border of the band}
 \end{aligned} \tag{4}$$

where the weight  $w_i$  is used to consider the normal estimation error. This is a positive definite quadratic optimisation problem with linear inequality constraints. We employ an active set block pivoting method [10] using a conjugate gradient solver with a limited memory incomplete Cholesky factorization [11] as preconditioner to take advantage of the sparsity of the problem. The regularization term utilized in the SSD penalizes curvature. In contrast, the

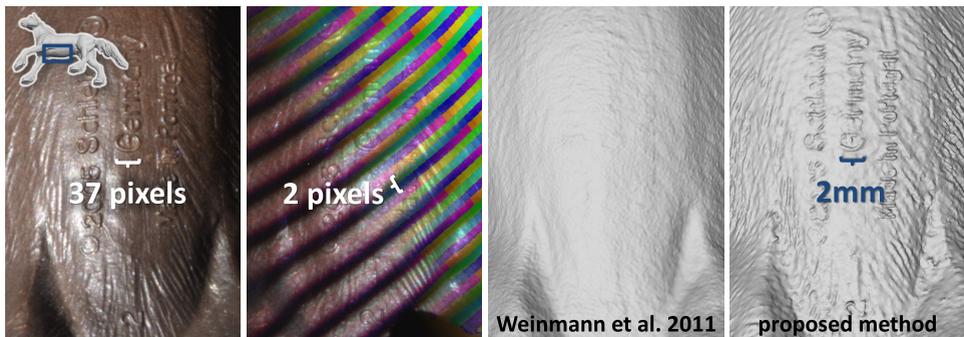


Figure 5: **A detail view of a toy horse** The first image is taken from the input data of the Helmholtz normal estimation and illustrates the resolution of the cameras. The second picture shows one of the structured light images overlaid with the decoded codewords and demonstrates the resolution of the projector. The last two pictures compare the reconstruction results obtained using the method proposed in [51] (left) and our method (right). The reconstructed writing exhibits a depth of approximately  $150 \mu\text{m}$ .

min-cut optimisation employs a minimum surface regularization. It is non-trivial to use the same regularization in both cases, due to fact that two different optimisation techniques are used. However, due to the fact that the signed distance function is restricted to a narrow band around the min-cut solution, a strong deviation of the smooth signed distance result from the min-cut solution is prevented. The final step in our reconstruction consists of converting the refined implicit function into a mesh. This is achieved by applying the isosurface extraction on octree grids presented in [47].

## 6 Results

We evaluate our technique both on a synthetic reference dataset as well as real world examples. All shown test cases, including the synthetic ones, were created for  $15^\circ$  steps of the turntable rotation and contain for each rotation a full measurement from our setup with 11 cameras, 4 projectors, and 198 lights.

In Figure 3, we show our results on the synthetic test dataset. For this experiment, we rendered images under a glossy reciprocal Phong BRDF and simulated projector-codeword images of the Lucy mesh [49]. A comparison with a purely structured light based reconstruction exploiting projector super-resolution [51] is shown. This technique first reconstructs a point cloud and then utilizes Poisson surface reconstruction [46] to obtain a mesh.

In Figure 4, we show a reconstruction of a real-world brass figurine. This demonstrates that our approach can cope with glossy material and reconstruct even fine surface details, which are lost in the reference laser scan and the structured light reconstruction due to noise in the triangulated points. Figure 5 illustrates, that even very small details far below the resolution of the projector patterns like the engraved text can be reconstructed.

All reconstructions were performed on a level 10 octree ( $1024^3$  nodes on the finest level). Using a parallelized C++ implementation, the computation of the Lucy dataset took about 8 hours on a computer with two Intel Xeon E5620 processors and 24GB of RAM. However, depending on the octree occupancy and the selected smoothing strength, timings

might vary from case to case.

## 7 Conclusions and Future Work

We demonstrated that the combination of structured light scanning with Helmholtz normal estimation enables the reconstructions of high-quality 3D models, faithfully representing even fine surface details. By applying a variational approach, we formulate the 3D reconstruction as one combined optimisation problem over all available input data.

Usually, it is not possible to get a full scan of an object without capturing at least two poses. These have to cover the upper and lower hemisphere and need to be combined. In the future, instead of merging these individual scans after the 3D reconstruction, this process could be integrated into the optimisation to get one globally consistent reconstruction.

Especially for very specular materials, the results could be further improved by adding outlier robust statistics to the Helmholtz normal estimation. For very challenging cases, these could also be complemented with alternative normal estimation techniques like shape-from-specularity to handle an even wider range of materials. We would also like to integrate additional consistency terms into the optimisation, such as colour consistency.

Finally, the possibility of performing reconstructions for large objects at high resolutions is currently limited by the time needed for the graph-cut and normal computations. Both algorithms are candidates for GPU implementations. However, this will require techniques to break the problem into smaller meaningful subproblems, due to limited memory of GPUs.

## Acknowledgements

The research leading to these results was partially funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231809; 2008-2012 and the German Science Foundation (DFG) under research grant KL 1142/4-2.

## References

- [1] M. Adlers. *Topics in Sparse Least Squares Problems*. PhD thesis, 2000.
- [2] D. G. Aliaga and Y. Xu. A self-calibrating method for photogeometric acquisition of 3D objects. *PAMI*, 32(4):747–754, 2010.
- [3] N. Birkbeck, D. Cobzas, M. Jagersand, and P. Sturm. Variational shape and reflectance estimation under changing light and viewpoints. In *ECCV*, pages 536–549, 2006.
- [4] F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1):231–243, 2004.
- [5] F. Calakli and G. Taubin. Ssd: Smooth signed distance surface reconstruction. *Comput. Graph. Forum*, 30(7):1993–2002, 2011.
- [6] D. Cremers and K. Kolev. Multiview stereo and silhouette consistency via convex functionals over convex domains. *PAMI*, 33(6):1161–1174, 2011.

- [7] A. Delaunoy and E. Prados. Gradient flows for optimizing triangular mesh-based surfaces: Applications to 3d reconstruction problems dealing with visibility. *IJCV*, 2011.
- [8] A. Delaunoy, E. Prados, and P. N. Belhumeur. Towards full 3D Helmholtz stereovision algorithms. In *ACCV 2010 - Volume Part I*, pages 39–52, 2011.
- [9] C. Hernández Esteban and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Comput. Vis. Image Underst.*, 96:367–392, 2004.
- [10] C. Hernandez Esteban, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *TPAMI*, 30(3):548–554, 2008.
- [11] S. Herbot and C. Wöhler. An introduction to image-based 3d surface reconstruction and a survey of photometric stereo methods. *3D Research*, 2(3):40:1–40:17, 2011.
- [12] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi. A hand-held photometric stereo camera for 3-d modeling. In *ICCV*, pages 1234–1241, 2009.
- [13] M. Holroyd, J. Lawrence, and T. Zickler. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Trans. Graph.*, 2010.
- [14] A. Hornung and L. Kobbelt. Hierarchical Volumetric Multi-view Stereo Reconstruction of Manifold Surfaces based on Dual Graph Embedding. In *CVPR*, pages 503–510, 2006.
- [15] H. Jin, D. Cremers, A.J. Yezzi, and S. Soatto. Shedding light on stereoscopic segmentation. In *CVPR*, volume 1, pages I–36 – I–42, 2004.
- [16] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *SGP*, pages 61–70, 2006.
- [17] M. Kazhdan, A. Klein, K. Dalal, and H. Hoppe. Unconstrained isosurface extraction on arbitrary octrees. In *Eurographics Symposium on Geometry Processing (SGP)*, pages 125–133, 2007.
- [18] K. Kolev, T. Pock, and D. Cremers. Anisotropic minimal surfaces integrating photo-consistency and normal information for multiview stereo. In *ECCV*, pages 538–551, 2010.
- [19] K. Kolev, N. Kirchgeißner, S. Houben, A. Csiszár, W. Rubner, C. Palm, B. Eiben, R. Merkel, and D. Cremers. A variational approach to vesicle membrane reconstruction from fluorescence imaging. *Pattern Recogn.*, 44(12):2944–2958, 2011.
- [20] A. Ladikos, S. Benhimane, and N. Navab. Multi-view reconstruction using narrow-band graph-cuts and surface normal optimization. In *BMVC*, pages 15.1–15.10, 2008.
- [21] C.-J. Lin and J. J. Moré. Incomplete Cholesky Factorizations with Limited Memory. *SIAM J. Sci. Comput.*, 21(1):24–45, 1999.
- [22] Z. Lu, Y.-W. Tai, M. Ben-Ezra, and M. S. Brown. A framework for ultra high resolution 3d imaging. In *CVPR*, 2010.
- [23] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM Trans. Graph.*, 24:536–543, 2005.

- [24] M. Ritz, M. Scholz, M. Goesele, and A. Stork. High resolution acquisition of detailed surfaces with lens-shifted structured light. In *VAST*, pages 1–8, 2010.
- [25] G. Sansoni, M. Trebeschi, and F. Docchio. State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, 9(1):568–601, 2009.
- [26] C. Schwartz, M. Weinmann, R. Ruiters, and R. Klein. Integrated high-quality acquisition of geometry and appearance for cultural heritage. In *VAST*, pages 25–32, 2011.
- [27] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, volume 1, pages 519–528, 2006.
- [28] S. N. Sinha and M. Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *ICCV*, volume 1, pages 349–356, 2005.
- [29] The Stanford 3D Scanning Repository. <http://graphics.stanford.edu/data/3Dscanrep/>.
- [30] G. Vogiatzis, C. Hernández, P. H. S. Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts and occlusion robust photo-consistency. *PAMI*, 29(12):2241–2246, 2007.
- [31] M. Weinmann, C. Schwartz, R. Ruiters, and R. Klein. A multi-camera, multi-projector super-resolution framework for structured light. In *3DIMPVT*, pages 397–404, 2011.
- [32] 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.
- [33] Y. Yoshiyasu and N. Yamazaki. Topology-adaptive multi-view photometric stereo. In *CVPR*, pages 1001–1008, 2011.
- [34] T. Yu, N. Ahuja, and W. C. Chen. Sdg cut: 3d reconstruction of non-lambertian objects using graph cuts on surface distance grid. In *CVPR*, pages 2269–2276, 2006.
- [35] J. Yuan, E. Bae, and X.-C. Tai. A study on continuous max-flow and min-cut approaches. In *CVPR*, 2010.
- [36] T. Zickler, P. N. Bellhumeur, and D. J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. In *IJCV*, pages 869–884, 2002.