Combining Photometric Normals and Multi-View Stereo for 3D Reconstruction

Martin Grochulla
MPI Informatik
Universität Campus
Saarbrücken, Germany
mgrochul@mpi-inf.mpg.de

Thorsten Thormählen
Philipps-Universität Marburg
Hans-Meerwein-Strasse
Marburg, Germany
thormae@informatik.uni-marburg.de

ABSTRACT

In this paper a novel approach for 3D reconstruction of mid-sized objects is proposed that combines the advantages of multi-view stereo and photometric normal estimation. Images of an inspected object are taken in parallel with multiple synchronized cameras while the object is placed in an illumination setup that produces time-varying spherical gradient illumination for normal estimation. In contrast to existing approaches, the normal information is not only used to refine the preliminary stereo reconstruction. Instead, it is shown that the normal information allows to significantly increase the window size for stereo matching, which strongly improves robustness compared to classical window matching in the image domain. As a consequence, smoothness constraints, which are typical for stereo reconstructions, are no longer required. Our proposed method is comparably simple to implement and has a computation time of a few minutes. The evaluation on real and synthetic data demonstrates that detailed reconstructions with high accuracy are obtained.

Categories and Subject Descriptors

I4.5 [**Image Processing and Computer Vision**]: Reconstruction; I4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Shape, Texture*

General Terms

Algorithms

1. INTRODUCTION

3D reconstructions of mid-sized real objects are often obtained by triangulation methods. In this area one can differentiate between active and passive methods.

Active triangulation methods, such as laser scanning [18] or structured-light-scanning [23], modify the captured scene by projecting a high-frequency time-varying illumination pattern onto the object. The accuracy is typically limited by the resolution of the projected patterns. These methods are currently used the most in practical applications, because of their high robustness. However, the capturing speed is limited by the projector and acquisition hardware. Furthermore, for a complete reconstruction the object must be captured from multiple viewpoints. Because of interferences of the illumination patterns, acquisition from multiple viewpoints cannot be performed in parallel, which makes the overall process slow.

In contrast, passive triangulation methods, such as stereo [25] or multi-view-stereo [26], rely solely on two or more images taken from different viewpoints. The robustness of passive stereo strongly depends on matchable image features. Therefore, a passive stereo reconstruction requires a surface reconstruction method to obtain a closed surface. The main advantages of passive triangulation methods are their reduced hardware effort because no projectors are required and their speed because multiple viewpoints can be captured in parallel.

Both, passive and active triangulation methods, can be combined with photometric stereo. Photometric stereo [29, 20] is a technique that allows reconstructing high-resolution surface normals by observing the shading of an object under different lighting conditions from a single viewpoint. In contrast to active triangulation methods, the illuminations are not high-frequency patterns but rather multiple distributed low-frequency illuminations that allow to invert the shading model in order to estimate the surface normal.

In this paper we propose a novel, flexible method that can be classified as a combination of passive triangulation and photometric stereo. We generate a time-varying spherical gradient illumination and observe the scene with multiple cameras. Detailed normal maps of the inspected object from different viewpoints are obtained using photometric stereo. The normal maps are the input to our multi-view stereo algorithm which generates a 3D reconstruction. Then, the normal information is employed to interpolate between the sparsely sampled stereo estimates in order to obtain a high-resolution reconstruction.

Contribution. In contrast to most existing approaches, we employ the normal information not solely to refine the reconstruction. Instead, the normal information is used in the matching process of our multi-view stereo approach. Our approach enables us to use significantly larger windows during patch matching, which strongly increases its robustness. As a consequence, smoothness constraints that are typically required in passive stereo are not necessary. Consequently, without smoothness constraints the surface can be sparsely evaluated which vastly reduces the computational effort compared to a densely sampled evaluation. The resulting reconstruction is sparse, but the included estimates are all reliable.

Finally, the sparse reconstruction is interpolated with the normal information to obtain a dense and detailed reconstruction.

Limitation. Our approach is designed to handle and reconstruct objects, that are static, smooth and diffuse. The limitation to static objects results from the limited capturing speed of the employed consumer digital single lens reflex (DSLR) cameras. We focus on diffuse objects or objects that are best approximated as diffuse. However, specular objects can be reconstructed to a certain extent as presented in the results giving an indication about the robustness of our method.

2. RELATED WORK

This section reviews related work that also combines passive triangulation with photometric stereo for accurate 3D reconstruction.

Photometric Refinement. Many approaches have been proposed using photometric information to improve and refine an initial geometry or surface. The initial geometry in these approaches can be obtained either by (multi-view) stereo reconstruction [9, 2, 30, 31, 22], structure-from-motion [33, 19, 16, 24], or triangulation scanning [21, 17]. Other approaches employ 3D models that are morphed [32] or estimate shadow maps that are then used to reconstruct the 3D geometry [7].

In contrast to our approach, these methods have in common that normal information is not an integral part of the 3D position acquisition. Photometric information is used in a refinement step, but it is not directly employed for the generation of the initial 3D reconstruction.

Silhouettes-Based Approaches. Silhouette information extracted from multiple views allows to generate a visual hull of the object. The 3D positions and normals of the visual hull can be optimized to obtain a 3D model [6, 8, 15]. The visual hull can also be used as a proxy to deform and assemble partial reconstructions to a complete 3D model [27].

However, our approach does not rely on silhouette information and also works in situations where the visual hull is not available or is not very descriptive (for instance, for a frontal view of a relief).

Multi-view Stereo and Normals. Surface normals and positions can also be conjointly estimated using a set of images captured under multiple point light illuminations [4] by using a known example object [1].

In contrast to multi-view stereo methods our approach does not rely on detectable image features which lead to sparse point clouds where a surface has to be fitted.

Uncalibrated Photometric Stereo. Furthermore, uncalibrated photometric stereo refers to the case in which the lightning conditions are not known. Different methods deal with this scenario. They generate a 3D model of the object and can estimate the light positions [3, 11] or compensate for varying unknown illumination conditions [12].

3. ACQUISITION SETUP

In order to perform a 3D reconstruction with our method, the inspected object is placed in a special hardware setup that can generate time-varying illuminations. Images of the object under different illuminations are taken by multiple calibrated and synchronized DSLR cameras. In the following the illumination hardware and camera setup is presented.

3.1 Illumination Hardware and Photometric Stereo

Our illumination hardware is similar to the one made popular by Devebec and colleagues [20]. It is a metal frame in the shape of a sphere with a diameter of 150 cm. 160 white LEDs are evenly distributed and attached to the frame. Their brightness can be controlled individually to have full control of the lighting conditions inside the sphere.

The object is captured under six different gradient illuminations, which are axis parallel, resulting in a set $\mathcal{L}$ of six luminance images

$$\mathcal{L} = \left\{ \mathrm{L}^x, \mathrm{L}^{-x}, \mathrm{L}^y, \mathrm{L}^{-y}, \mathrm{L}^z, \mathrm{L}^{-z} \right\}. \tag{1}$$

Given a diffusely reflecting surface of an object that has been captured under these six illuminations, the normal map $\mathrm{N}(\cdot)$ is computed pixel-wise as proposed by Wilson et al. [28]:

$$\mathrm{N} = \frac{(\mathrm{L}^x - \mathrm{L}^{-x}, \mathrm{L}^y - \mathrm{L}^{-y}, \mathrm{L}^z - \mathrm{L}^{-z})^\top}{||(\mathrm{L}^x - \mathrm{L}^{-x}, \mathrm{L}^y - \mathrm{L}^{-y}, \mathrm{L}^z - \mathrm{L}^{-z})^\top||}. \tag{2}$$

Figure 1 shows the illumination hardware and the generated normal maps for different viewpoints. The normals are consistent across the views, which is the requirement to employ the normal information as a matching score for multi-view stereo.

3.2 Multi-view Camera Calibration

An accurate (geometric) camera calibration is important, because calibration errors directly propagate into the 3D reconstruction. The camera calibration estimates the relation between cameras by estimating the camera parameters. We use an approach similar to [13].

For the calibration we use an object that consists of several planar calibration patterns (see Figure 1). The calibration object is captured in different positions and orientations simultaneously by all cameras. This leads to a better coverage of the sampling volume and improves the quality of the estimated camera parameters. Additionally, one picture is acquired by each camera where the calibration object is aligned to the coordinate frame of the illumination hardware. This establishes the geometric relation between the illumination hardware and the cameras.

4. PHOTOMETRIC MULTI-VIEW STEREO

In this section, we present our method for multiple view 3D reconstruction using normal maps obtained from photometric stereo.

Overview. Low frequency noise that is present in the input normal maps renders direct integration methods unsuitable to accurately reconstruct the true 3D geometry of the object [21], while high frequency noise leads to wrong reconstructions of local surface features.

Instead, in the first step (detailed in Section 4.1), normal information from photometric stereo is used to improve the patch matching capabilities of multi-view stereo. In this step, 3D patch surfaces are generated using normal information in a reference view. Reliable and accurate depth values are obtained by optimizing the 3D

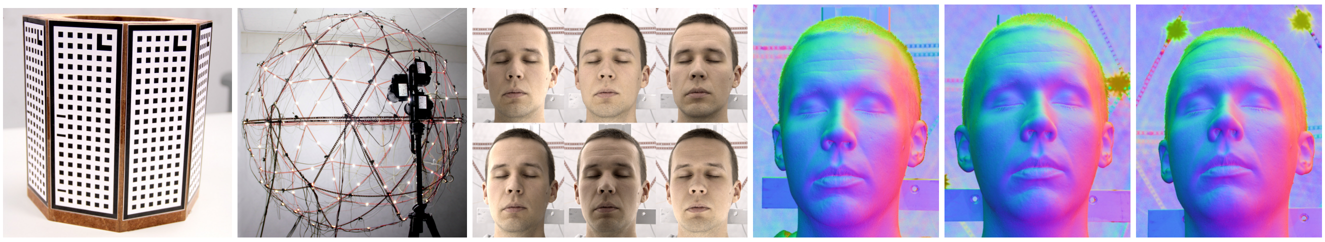

Figure 1: From left to right: Calibration object; Illumination hardware that generates different illuminations; Example input images for one viewpoint (all six spherical gradient illumination are shown); Generated normal maps (color-coded as RGB) for three different viewpoints.

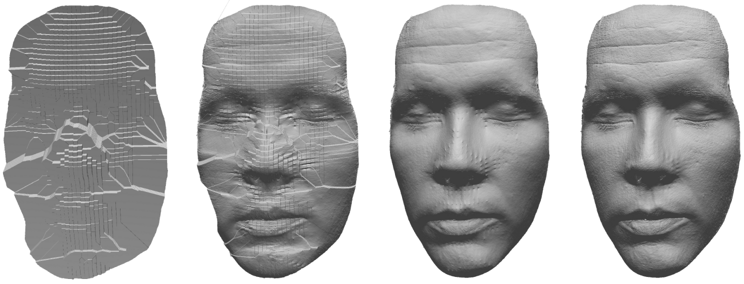

Figure 2: Overview of the different steps of the algorithm. From left to right: Initial sparse reconstruction after nearest neighbor interpolation; reconstruction obtained after 1 iteration of nonlinear least squares minimization; after 20 iterations of minimization; after 20 filtering iterations.

patch matching costs. However, computing the depth values for all pixels is computationally very expensive. Hence, we reconstruct the depths of a small number of points resulting in a sparse reconstruction. This reduces the computation time to a few minutes on standard PC hardware.

In the second step (detailed in Section 4.2), we use the normal map of the reference view to interpolate the sparse reconstruction keeping the sparse points fixed. This resulting reconstruction is dense and has high accuracy, however, there are some tiny disturbing peaks at the positions of the sparse 3D points because of the quantization along the line of sight.

In a third step (detailed in Section 4.3) those peaks are removed by a normal map driven filtering. Figure 2 shows the output of the individual steps of our algorithm for a face. The next section explains each step in more detail.

It should be stressed that the normal map driven filtering is different from introducing a smoothness term in the dense reconstruction step. In the reconstruction step we want to compute a dense surface that approximates the measured normal maps the best. Introducing a smoothness term results in smoothing the overall reconstructed surface, which is not desirable. In contrast, the normal map driven filtering is able to smooth areas where normal map and surface contradict each other, while enhancing areas, where normal map and reconstructed surface agree.

4.1 Sparse reconstruction

In this first step an initial point cloud of the object is reconstructed that consists of reliable and accurate 3D points. For the initialization a reference view v_{ref} is set. The 3D points of the initial point cloud $\mathcal{I}$ correspond to 2D points lying on an equidistant grid in the reference view and their depths. For each grid point we use the normals given in a window of size $w \times w$ to reconstruct the local 3D geometry of that patch surface as follows.

Patch Reconstruction. Given the lines of sight $\mathbf{l}_i$ and $\mathbf{l}_j$ of pixels i and j, the normal $\mathbf{n}_i$ and the candidate depth d_i of pixel i. We compute the depth d_j of pixel j as the intersection of the ray with direction $\mathbf{l}_j$ and the plane with normal $\mathbf{n}_i$ located at $d_i \cdot \mathbf{l}_i$ by

$$d_j = \frac{\mathbf{l}_i \bullet \mathbf{n}_i}{\mathbf{l}_j \bullet \mathbf{n}_i} \cdot d_i, \tag{3}$$

where • denotes the scalar product (see Figure 3). Rewriting this equation yields the constraint

$$\mathbf{l}_j \bullet \mathbf{n}_i \cdot d_j - \mathbf{l}_i \bullet \mathbf{n}_i \cdot d_i = 0. \tag{4}$$

Stacking up equations for all neighbors in a 4-neighborhood of every pixel in the considered window leads to a linear system of equations $\mathtt{A}' \cdot \mathbf{d} = \mathbf{0}$, where $\mathbf{d}$ is the vector of unknown depths. Additionally, every equation (i.e., row of matrix $\mathtt{A}'$) is weighted according to a standard distribution with standard deviation σ depending on the distance of the considered pixel to the center pixel of the window.

In order to avoid the trivial solution we set the depth of the center pixel to 1 leading to the extended linear system of equations: $\binom{\mathtt{A}'}{\mathbf{e}^\top} \cdot \mathbf{d} = \binom{\mathbf{0}}{1}$, where $\mathbf{e}$ is the canonical unit vector corresponding to the position of the center pixel in vector $\mathbf{d}$. Choosing a different depth d_i for the center pixel i will lead to $d_i \cdot \mathbf{d}$ as solution for the overdetermined linear system of equations. Writing the extended linear system of equations as

$$\mathtt{A} \cdot \mathbf{d} = \mathbf{b}, \tag{5}$$

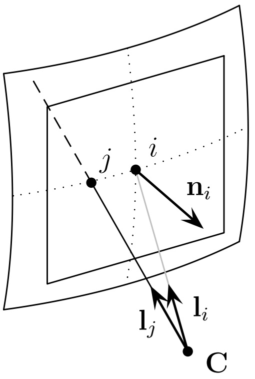

Figure 3: Computation of depths. Given the normal $\mathbf{n}_i$ of pixel i with corresponding line of sight $\mathbf{l}_i$, which is at depth d_i. The depth d_j of the neighbor pixel j with corresponding line of sight $\mathbf{l}_j$ is computed as ray-plane intersection as given in Eq. 3.

we compute a least squares solution by solving $\mathtt{A}^\top \cdot \mathtt{A} \cdot \mathbf{d} = \mathtt{A}^\top \cdot \mathbf{b}$. The point cloud $\mathcal{P}_i$ representing the local 3D surface of that patch for pixel i is obtained by multiplying the depths with their corresponding line of sights:

$$\mathcal{P}_i := \left\{ \mathbf{X}_j | \mathbf{X}_j = \mathbf{l}_j \cdot d_j, 1 \leqslant j \leqslant w^2 \right\}. \qquad (6)$$

Line of Sight Sampling. In order to find the 3D point $\mathbf{X}_i$ for each 2D grid pixel i, we sample depths d_i on the line of sight. We can simply transform the reconstructed patch $\mathcal{P}_i$ using $d_i \cdot \mathbf{d}$. This is a crucial advantage of our formulation because re-solving Eq. 5 is not required for each depth candidate. Otherwise, the sampling along the line of sight would be computationally too expensive for large patch size.

We project the reconstructed patch $\mathcal{P}_i$ from Eq. 6 at depth d_i into the other views $v \neq v_{\text{ref}}$. We then compute the 3D patch matching cost $\mathrm{c}(d_i)$ for depth d_i by comparing the projections of the normals in the window around the grid point with the normals at the points of projection in the other views v:

$$\mathrm{c}(d_i) = \sum_{v \neq v_{\text{ref}}} \sum_{\mathbf{X}_j \in \mathcal{P}_i} \mathrm{d}(\mathrm{N}_{v_{\text{ref}}}(\mathtt{P}_{v_{\text{ref}}} \mathbf{X}_j \cdot d_i), \mathrm{N}_v(\mathtt{P}_v \mathbf{X}_j \cdot d_i))^2, \qquad (7)$$

where $\mathbf{X}_j$ is the j-th 3D point obtained from patch reconstruction, $\mathtt{P}_v$ is the projection matrix of camera v, and $\mathrm{N}_v(\cdot)$ denotes the normal map of view v, returning the interpolated value at the given point. We then choose the depth d_i that has the lowest cost among all depths.

A typical window size for local patch reconstruction is 160×160 pixels, which is much larger than patch sizes used by standard multi-view stereo in the image domain (which is typically 16×16 pixels or less). Increasing the patch size in the image domain is not possible because the true surface can no longer be approximated by a fronto-parallel plane as assumed by standard 2D patch matching. Fig 4 shows a comparison of standard 2D patch matching and our matching via local 3D patch surfaces. Because a large patch can contain much information, the cost function in Eq. 7 typically has a single distinct minimum. However, if this is not the case, we discard estimates that have only small variations among similar depths:

$$\mathrm{Var}\left[\mathrm{c}(d_i - D_{\mathrm{Var}}, d_i + D_{\mathrm{Var}})\right] < t_{\mathrm{Var}}. \qquad (8)$$

As a consequence, our approach is highly robust. In our experiments, we observed that the reconstructed depths contain no outliers if the threshold t_{Var} was chosen correctly. As a result of this initialization step we obtain a set $\mathcal{I}$ of grid points in the reference view v_{ref} with corresponding depths.

4.2 Dense Reconstruction by Nonlinear Optimization

In the second step, the set of points $\mathcal{I}$ with depths d_i from the multi-view stereo approach and the normal map $\mathrm{N}_{v_{\text{ref}}}$ is used to reconstruct the dense surface of the object. This is done by minimizing the cost function

$$\bar{c}(\mathbf{d}) = \sum_k \sum_{j \in \mathcal{N}(k)} \left(\mathbf{l}_j \bullet \mathbf{n}_i \cdot d_j - \mathbf{l}_k \bullet \mathbf{n}_k \cdot d_k \right)^2$$
$$\text{subject to } d_k = d_i,\ \forall i \in \mathcal{I}, \qquad (9)$$

where $\mathcal{N}(k)$ denotes the 4-neighborhood of pixel k. Again, $\mathbf{d}$ is the vector of unknown depth, now containing all depths d_k. This cost function penalizes deviations from Eq. 4, while fixing the set of 3D points $\mathcal{I}$. This is done to prevent the normals from pulling the reconstruction towards the unconstrained solution, which is known to exhibit low-frequency errors [21] on one hand and to avoid heading towards the trivial solution on the other hand. A user-defined binary mask is used to determine the region of interest which determines the set of pixels k used in the reconstruction process.

The cost function Eq. 9 is minimized using non-linear least squares: As initial solution we use the 3D point cloud $\mathcal{I}$ of the grid points with depths from the first step. Values in between the grid are interpolated by nearest neighbor. The cost function is assumed to be locally linear and is iteratively minimized. In each iteration a linear least squares problem similar to Eq. 5 is solved. In all experiments 20 iterations have been performed.

4.3 Filtering

In the last step, we filter the depth values obtained from minimizing Eq. 9. This is because the depths d_i of the set of initial 3D points $\mathcal{I}$ have not been optimized in order to avoid the trivial solution. Filtering the depth values is done iteratively. Rewriting Eq. 4 we obtain

$$d_j = \frac{\mathbf{l}_i \bullet \mathbf{n}_i}{\mathbf{l}_j \bullet \mathbf{n}_i} \cdot d_i, \qquad (10)$$

which is used to propagate the four depth values of the four neighbors (left, right, top, and bottom neighbor) to pixel i.The average depth of the four propagated depth values is computed and updates the depth of pixel i. These depth values geometrically correspond to line-plane-intersections. For numerical stability, a propagated depth value is only used for the update, if the angle between the line of sight $\mathbf{l}_i$ and the normal $\mathbf{n}_j$ is greater than $\arccos(t_{\text{angle}})$ with threshold parameter t_{angle}. In our experiments we use $t_{\text{angle}} = 0.173$ corresponding to an angle of approximately 80°. This update is performed iteratively for all pixels. In all experiments we performed 20 iterations.

5. RESULTS

In this section we evaluate our method on synthetic data, demonstrate 3D reconstructions of real objects, compare it to laser scanning.

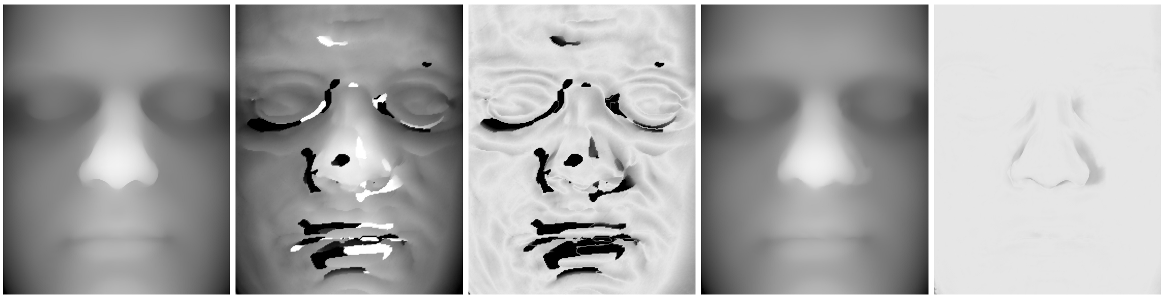

Figure 4: Comparison of patch matching with synthetic head model from three views, patches of size 32×32 pixels used. From left to right: ground truth depth map; depth map obtained from 2D patch matching; difference between ground truth and obtained depth map (gray values adjusted for better visibility); depth map obtained from 3D patch matching; difference between ground truth and obtained depth map (gray values adjusted).

5.1 Synthetic Data

Our method is evaluated on synthetic data. We generate a series of synthetic images of a 3D model of a human head as ground truth. We compute the normal maps for all three views. The three normal maps and the positions of the cameras are used to reconstruct the human head. We align the ground truth head model and our reconstructions with the *iterative closest point* (ICP, [5]) algorithm. The average distance between the mesh of the head model and the one of our reconstruction is used as quality measure. We test our method in settings with different material properties: Lambertian reflectance without shadows; Lambertian reflectance with shadows and different levels of Gaussian noise added to the input images; Lambertian reflectance with shadows, specular reflections, and different levels of Gaussian noise. The results are shown in Table 1: The best reconstruction is obtained in the ideal setting. With increasing level of noise the quality of the reconstruction decreases, while in general the head is reconstructed better in the absence of specular reflections (because specular reflections violate the assumption of a perfectly Lambertian surface).

5.2 Real-World Data

For demonstrating the applicability of our method we show several reconstructions of real-world objects. We use three digital single lens reflex (DSLR) cameras. Six images of the objects under the gradient illuminations and one additional image with uniform illumination are acquired. The resolution of the input images is 2592×1728 pixels. Figure 5 shows results of our reconstruction method for five objects: Relief, Purse, Shoe, Santa, and Vase. Additional close-up views of the 3D geometry with and without texture of Relief, Shoe, and Vase examples are shown in Figure 6. Although the Vase has a glossy surface, strong errors in the reconstruction are mainly visible at grazing angles, while other parts are reconstructed fairly well. This demonstrates that our approach is able to handle deviations from the assumptions to some extent. Figure 7 shows the results of reconstructing three faces. For all reconstructions we used a window size of 160×160 pixels, 20 iterations for minimizing Eq. 9, and 20 additional iterations for filtering as described in Section 4.3. t_{Var} has been set to 30.0. On standard PC hardware using unoptimized C++ code, the computation times are between 10 and 15 minutes depending on the model.

5.3 Comparison to Laser Scanning

We compare our method to 3D reconstructions of a laser scanner. Figure 8 shows a qualitative comparison between the scanned objects and our results. Our method is able to reconstruct finer details of the objects' surfaces. We align our 3D reconstructions and the ones obtained from laser scanning with the ICP method for a quantitative comparison. The distance between the two aligned meshes is used as a measure of quality. We obtain an average alignment error of 0.539% for the object Relief, 0.775% for the object Shoe, and 0.544% for the object Santa. The error is given relative to the size of the scanned object.

6. CONCLUSION AND FUTURE WORK

We have presented a novel approach that uses photometric normals in a multi-view stereo approach. The generated reconstructions are detailed and of high accuracy.

Currently, our results are obtained only from the perspective of a single reference view. Hence, they can be parametrized in the image domain and are in fact only 2.5D (2D plus depth). Augmenting our reconstructions to more complex topologies is possible by merging multiple 2.5D reconstructions. To do that, the number of cameras must be increased in order to capture multiple 2.5D reconstructions in parallel from different view points for increased surface coverage. Merging multiple 2.5D reconstructions to build a 3D mesh is out of the scope of this paper but a body of existing methods is available, such as the volumetric method of the volumetric range image processing package (VRIP, [10]).

Also, we do not handle cases in which the normal maps contain significant errors, for example, at depth discontinuities, in shadowed regions, or when the material is not diffuse and deviates from our assumption. This may lead to wrong reconstructions in the dense reconstruction step. This effect can be reduced by increasing the sampling density of the grid, having no disadvantage except for a larger computational effort. Furthermore, once a 3D mesh is recovered, shadowed regions can be detected (compare with [14]) and used to improve the generated normal maps.

In theory, our approach is capable of capturing dynamic scenes. However, the cameras are capable of only capturing approximately 4 images per second. This is rather low, considering that one data set in our approach consists of 7 images. Furthermore, the employed consumer DSLR cameras are difficult to synchronize, if operated at their maximum speed. In future, we would like to augment our capturing setup with more professional high-speed high-resolution cameras to obtain dynamic reconstructions.

	diffuse	diffuse with shadows				
Gaussian noise level [%]	0	0	1	2	3	4
AAE [%]	0.242	0.315	0.374	0.436	0.477	0.516
		diffuse and specular with shadows				
Gaussian noise level [%]		0	1	2	3	4
AAE [%]		0.402	0.434	0.468	0.528	0.544

Table 1: Average alignment error (AAE) of reconstructions to ground truth. The noise level is given as percentage of the range of pixel values of the input images. The error is given relative to the size of the ground truth model.

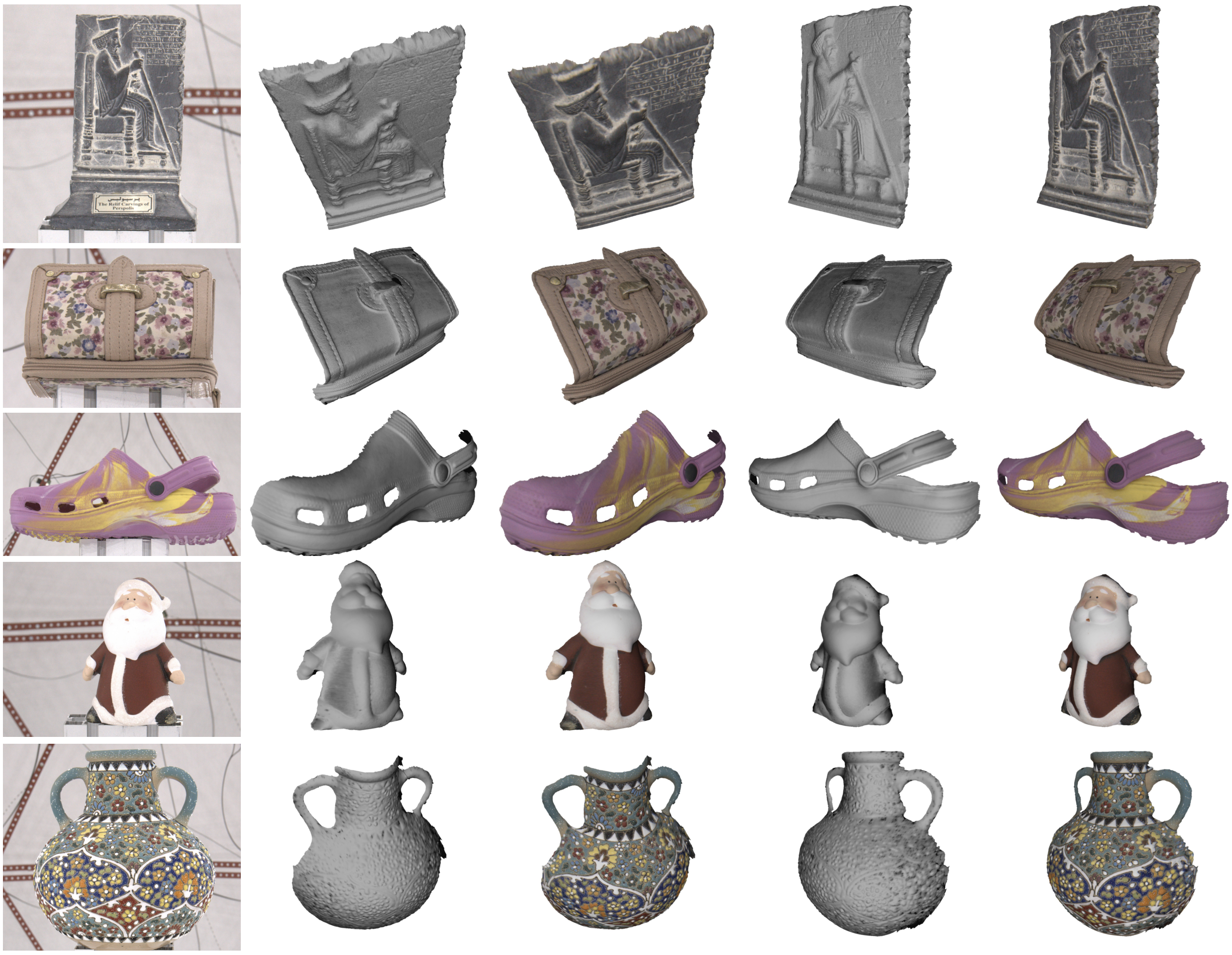

Figure 5: 3D Reconstructions of several objects. From top to bottom: Relief, Purse, Shoe, Santa, Vase. From left to right: Image from the reference camera; resulting 3D reconstruction shown from two different views with and without texture.

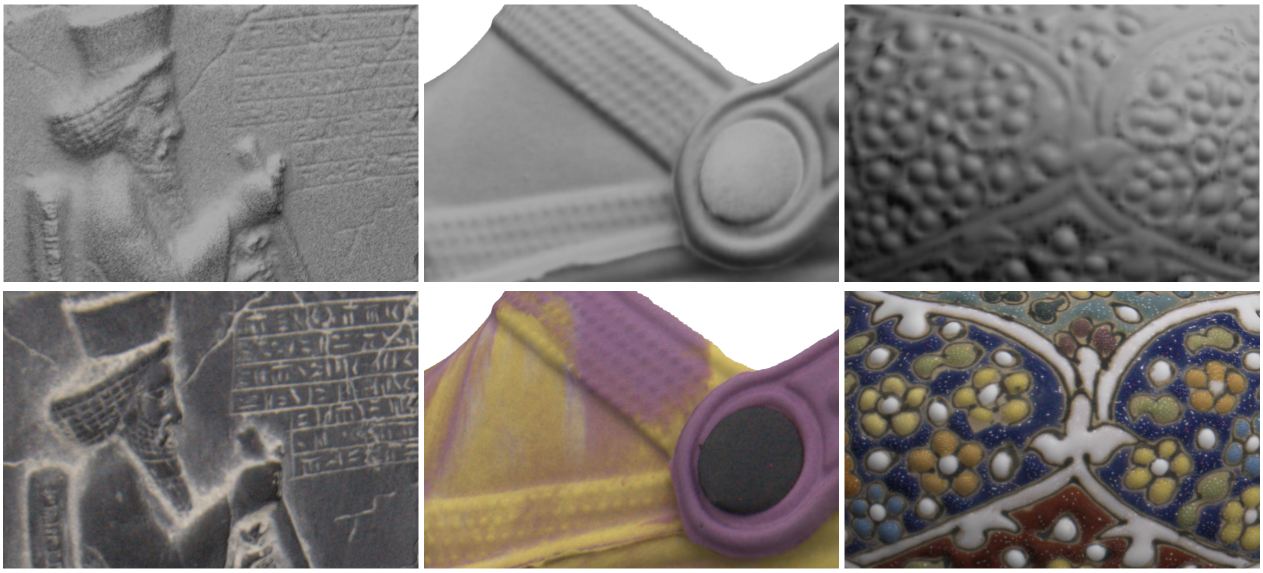

Figure 6: Details of the reconstructions. Top row; Reconstructed 3D geometry. Bottow row: Reconstructed 3D geometry with texture. From left to right: Relief, Shoe, Vase.

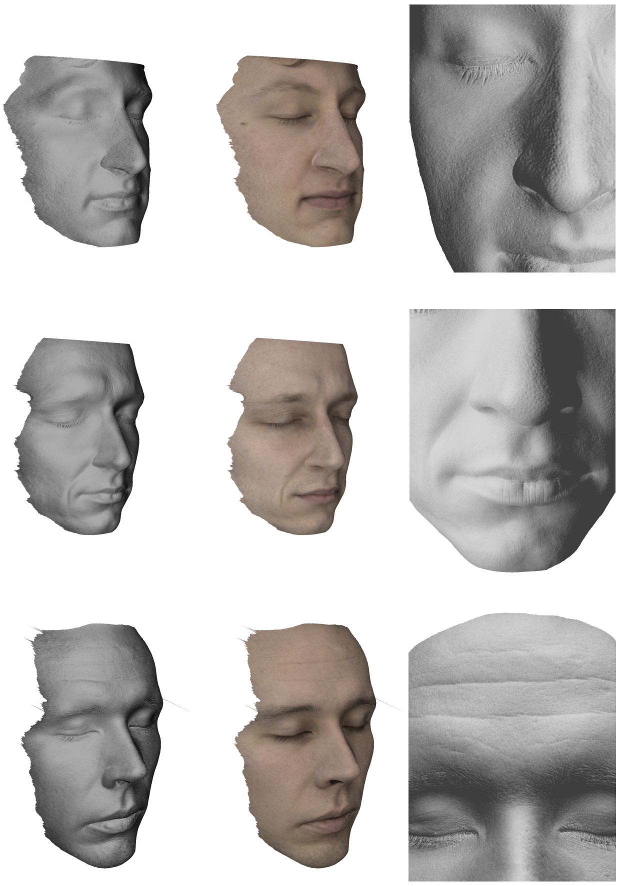

Figure 7: Reconstructions of three faces. From left to right: Reconstructed 3D geometry, rendered 3D reconstruction with texture, close-up view.

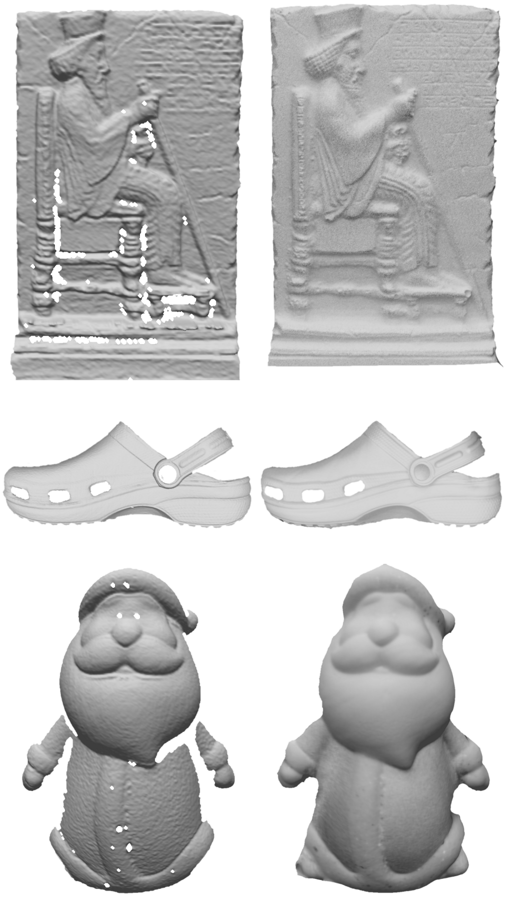

Figure 8: Comparison to laser scanning: Relief, Shoe, Santa. Left: results from laser scanner. Right: our reconstruction result.

7. REFERENCES

[1] J. Ackermann, F. Langguth, S. Fuhrmann, A. Kuijper, and M. Goesele. Multi-View Photometric Stereo by Example. In *International Conference on 3D Vision (3DV)*, 2014.

[2] R. Anderson, B. Stenger, and R. Cipolla. Color photometric stereo for multicolored surfaces. In *IEEE International Conference on Computer Vision*, pages 2182–2189, 2011.

[3] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *Int. J. Comput. Vision*, pages 239–257, 2007.

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

[5] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 239–256, 1992.

[6] N. Birkbeck, D. Cobzas, P. Sturm, and M. Jägersand. Variational shape and reflectance estimation under changing light and viewpoints. In *ECCV*, pages 536–549, 2006.

[7] M. K. Chandraker, S. Agarwal, and D. J. Kriegman. Shadowcuts: Photometric stereo with shadows. In *CVPR*, 2007.

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

[9] J. E. Cryer, P. Tsai, and M. Shah. Integration of shape from shading and stereo. *Pattern Recognition*, pages 1033–1043, 1995.

[10] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *SIGGRAPH*, pages 303–312, 1996.

[11] P. Favaro and T. Papadhimitri. A closed-form solution to uncalibrated photometric stereo via diffuse maxima. In *CVPR*, pages 821–828, 2012.

[12] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In *ICCV*, 2007.

[13] M. Grochulla, T. Thormählen, and H. Seidel. Using spatially distributed patterns for multiple view camera calibration. In *MIRAGE*, pages 110–121, 2011.

[14] D. C. Hauagge, S. Wehrwein, K. Bala, and N. Snavely. Photometric ambient occlusion. In *CVPR*, pages 2515–2522, 2013.

[15] C. Hernandez, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 548–554, 2008.

[16] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi. A hand-held photometric stereo camera for 3-d modeling. In *ICCV*, 2009.

[17] N. Joshi and D. J. Kriegman. Shape from varying illumination and viewpoint. In *ICCV*, 2007.

[18] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. E. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project: 3d scanning of large statues. In *SIGGRAPH*, pages 131–144, 2000.

[19] J. Lim, J. Ho, M. Yang, and D. J. Kriegman. Passive photometric stereo from motion. In *ICCV*, pages 1635–1642, 2005.

[20] W.-C. Ma, T. Hawkins, P. Peers, C. Chabert, M. Weiss, and P. E. Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Rendering Techniques*, pages 183–194, 2007.

[21] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM Trans. Graph.*, pages 536–543, 2005.

[22] J. Park, S. N. Sinha, Y. Matsushita, Y. Tai, and I. Kweon. Multiview photometric stereo using planar mesh parameterization. In *IEEE International Conference on Computer Vision*, pages 1161–1168, 2013.

[23] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, pages 2666–2680, 2010.

[24] P. Saponaro, S. Sorensen, S. Rhein, A. R. Mahoney, and C. Kambhamettu. Reconstruction of textureless regions using structure from motion and image-based interpolation. In *ICIP*, pages 1847–1851, 2014.

[25] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, pages 7–42, 2002.

[26] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, pages 519–528, 2006.

[27] D. Vlasic, P. Peers, I. Baran, P. E. Debevec, J. Popovic, S. Rusinkiewicz, and W. Matusik. Dynamic shape capture using multi-view photometric stereo. *ACM Trans. Graph.*, pages 1–11, 2009.

[28] C. A. Wilson, A. Ghosh, P. Peers, J. Chiang, J. Busch, and P. E. Debevec. Temporal upsampling of performance geometry using photometric alignment. *ACM Trans. Graph.*, 29(2), 2010.

[29] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):191139–191139–, 1980.

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

[31] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *CVPR*, pages 969–976, 2011.

[32] C. Yang, J. Chen, N. Su, and G. Su. Improving 3d face details based on normal map of hetero-source images. In *CVPR Workshops*, June 2014.

[33] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *ICCV*, pages 618–625, 2003.