Removing the Example from
Example-based Photometric Stereo

Jens Ackermann[1], Martin Ritz[2], André Stork[2], and Michael Goesele[1]

[1] TU Darmstadt
[2] Fraunhofer IGD

Abstract. We introduce an example-based photometric stereo approach that does not require explicit reference objects. Instead, we use a robust multi-view stereo technique to create a partial reconstruction of the scene which serves as scene-intrinsic reference geometry. Similar to the standard approach, we then transfer normals from reconstructed to unreconstructed regions based on robust photometric matching. In contrast to traditional reference objects, the scene-intrinsic reference geometry is neither noise free nor does it necessarily contain all possible normal directions for given materials. We therefore propose several modifications that allow us to reconstruct high quality normal maps. During integration, we combine both normal and positional information yielding high quality reconstructions. We show results on several datasets including an example based on data solely collected from the Internet.

1 Introduction

Passive large-scale geometry reconstruction of outdoor scenes has so far mostly relied on (multi-view) stereo techniques. In contrast, photometric stereo approaches have rarely been used on outdoor scenes—mostly due to the lack of control over the scene, illumination conditions, and capture setup (see Section 2 for details). In fact, we are not aware of any large scale photometric stereo approach. In this paper, we therefore take a step in this direction and propose a novel photometric stereo technique generalizing photometric stereo by example [1]. The approach is applicable to very general indoor and outdoor scenes and demonstrates strong improvements in terms of accuracy and completeness compared to standard multi-view stereo approaches.

Photometric stereo by example [1] is an elegant method to determine normal maps from a set of images with fixed viewpoint and varying, distant illumination. For each pixel, the vector of color values in all input images is matched to the closest vector of color values of pixels on one or more reference objects with known geometry. The corresponding normals are then transferred back yielding a complete normal map. Photometric stereo by example has two key advantages. First, lighting can be general and unknown and does not need to be reconstructed. Second, it works for objects with a broad range of reflectance properties as long as they are well approximated by the reference objects. Photometric stereo by example is therefore one of the most general photometric stereo techniques known today. There is, however, one disadvantage: Current techniques require explicit reference objects in the scene from which the normals are transferred. Scenes without reference object cannot be reconstructed.
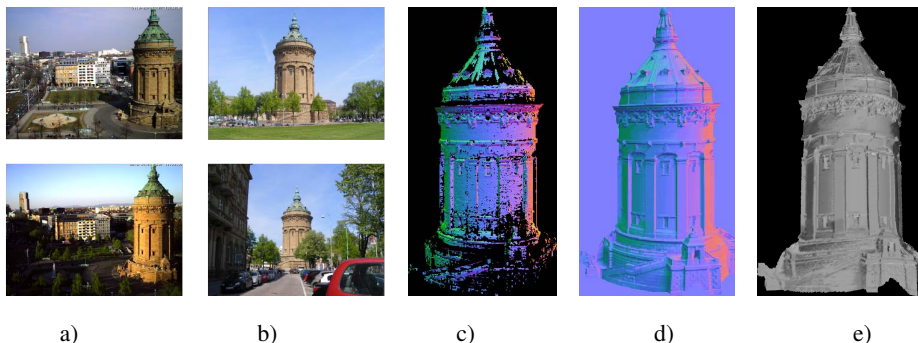
Fig. 1. *Tower* model. *a)* images captured by a static webcam for photometric stereo, *b)* images taken casually without special capturing setup for multi-view stereo, *c)* normal map of partially reconstructed geometry using multi-view stereo serving as scene-intrinsic reference geometry (SIRG), *d)* reconstructed normal map using photometric stereo with SIRG as reference object, *e)* final model rendered from novel viewpoint.

Our key observation is that many objects' geometry can at least be partially reconstructed using multi-view stereo as long as additional images are available that provide sufficient parallax (see, e.g., Figure 1). We propose to use this partial geometric scene model after suitable processing as *scene-intrinsic reference geometry* (SIRG) for a standard photometric stereo by example approach. This approach works if the reconstructed geometry (and therefore the normals) of the SIRG are sufficiently accurate and if the range of represented normals is wide enough to cover the normal directions represented in the scene. In addition, the reconstructed reference geometry should be a good representation of the reflectance properties in the scene, a condition that is often met since it is actually part of the scene.

Removing the need for explicit reference objects strongly extends the applicability of example-based photometric stereo at a comparably small acquisition cost (just a couple additional images for the multi-view stereo reconstruction). We demonstrate this with two examples where we base our reconstruction partially or completely on imagery available from Internet sources. The resulting normal maps and integrated geometry are nevertheless of high quality.

2   Related Work

Photometric stereo was introduced by Woodham [2] who assumed known distant point lighting and a known parametric reflectance model. Given three images of a diffuse surface from the same viewpoint, it is possible to determine the surface normal unless the illumination directions are coplanar. The basic theory of photometric stereo was then developed in the 1980s (see, e.g., Horn [3] for an overview) and research focused on generalizing it in various ways.

For example, Basri and Jacobs [4] introduced a system that simultaneously recovers unknown distant lighting. Illumination is estimated using a low degree spherical

harmonics basis suitable for approximately diffuse objects [5]. Shen and Tan [6] extended this technique to images with varying viewpoints but determine normals only at sparse points matched between the images. They demonstrate their approach also on images downloaded from an online image collection. Joshi et al. [7] propose a combination of multi-view and photometric stereo. They first reconstruct a rough geometry model using multi-view stereo and refine it with a photometric stereo approach. Hernandez et al. [8] describe a multi-view photometric stereo approach that additionally takes silhouette information into account. A similar approach was shown to work for dynamic scenes by Vlasic et al. [9] using a highly controlled capture setup. Higo et al. [10] introduce a system that simultaneously optimizes photoconsistency, normals, and surface smoothness.

Goldman et al. [11] use known lighting directions and cluster the surface in different materials. For each cluster, they determine the parameters of an analytic BRDF. Alldrin et al. [12] follow a similar approach but use a data-driven reflectance model instead of the analytic BRDF model.

2.1 Example-based Photometric Stereo

Based on Woodham's ideas, Silver [13] applied photometric stereo to objects with uniform but unknown surface reflectance. A matte white sphere serving as calibration object is captured under three different lighting conditions. Given its known geometry, one can construct a lookup table matching triples of intensity values with the sphere's surface gradient. For reconstruction, other matte white objects are captured under the same lighting conditions; corresponding surface gradients are determined using the lookup table.

Hertzmann and Seitz [1] generalized this approach using the orientation-consistency cue: Two points with the same surface orientation reflect the same light toward the viewer if they have the same BRDF, all light sources are distant, the camera is orthographic, and the points are not influenced by non-local lighting effects (e.g., shadows, interreflections). Their approach is very general and operates with arbitrary distant lighting on a very wide class of materials while still yielding high quality results. It requires, however, one or two reference objects in the captured scene that are used for normal transfer.

Koppal and Narasimhan [14] also exploit orientation-consistency to find clusters of iso-normals in a scene captured by a video camera. They do not require a reference object, but rely on a continuous, unstructured light source path and a dense sampling in the time domain. In an additional step, a classical photometric stereo approach can assign absolute normals to the clusters or other techniques can use the clusters as starting point for more detailed reconstructions.

In this work, we show that detailed normal maps can be reconstructed without the need for explicit reference objects or densely sampled video. We build on the standard photometric stereo by example approach [1] but replace the separate reference objects by the captured scene's own geometry which we partially and approximately reconstruct using a robust multi-view stereo technique.

3   Scene-Intrinsic Reference Geometry

Given a static scene, we capture multiple images $I^{\mathrm{PS}} = \{I_1^{\mathrm{PS}}, \ldots, I_n^{\mathrm{PS}}\}$ from the same camera position under unknown, distant, varying illumination for photometric stereo. We make the standard assumption that the camera well approximates an orthographic camera. We additionally capture another set of images $I^{\mathrm{MVS}} = \{I_1^{\mathrm{MVS}}, \ldots, I_m^{\mathrm{MVS}}\}$ from varying viewpoints for multi-view stereo. The latter images should provide sufficient parallax and lighting suitable for multi-view stereo reconstruction. The images I^{MVS} and one of the images for photometric stereo, without loss of generality I_1^{PS}, are registered using a robust structure from motion system [15]. Since the images I^{PS} were taken with identical intrinsic and extrinsic camera parameters, all images are now registered into a common coordinate system. Our goal is to first reconstruct a (partial) geometry model that serves as scene intrinsic reference geometry. Using the reference geometry, we then aim at creating a complete and accurate normal map. We finally reconstruct the scene geometry by integrating the resulting normal field while taking the reconstructed reference geometry into account.

There is a large body of existing work on multi-view stereo reconstruction (see Seitz et al. [16] and the accompanying web page) and our proposed technique can be based on any of them. Since we aim at handling very general input data, we selected the method of Goesele et al. [17] (see Section 6 for a comparison with a different algorithm) that is known to be robust and accurate even for very general input data. This method reconstructs individual, incomplete depth maps using a region-growing approach. We merge these depth maps into a combined triangular geometry model using volumetric range image processing (VRIP) [18]. This approach exploits redundancy in the input depth maps to reduce noise and remove outliers. It also assigns confidence values to vertices which we use to remove less reliable geometry from the reference geometry. Finally, we compute per-vertex normals for the reference geometry from surrounding face normals using area-weighted averaging. Using a variant of Laplacian smoothing, we iteratively smooth the computed normals according to

$$\mathbf{n}^k = \mathbf{n}^{k-1} + \lambda \sum_{i \in \mathcal{N}} \left(\frac{\mathbf{n}_i^{k-1} - \mathbf{n}^{k-1}}{|\mathcal{N}|} \right) \tag{1}$$

where $\mathcal{N}$ describes the neighborhood of $\mathbf{n}$. The resulting normal vector is normalized. In our standard matching, we perform 10 iterations with $\lambda = 0.05$.

4   Correspondence and Normal Transfer

In this section, we describe the details of our example-free photometric stereo by example approach. We first introduce the basic matching as in [1], restricted to a single reference object. In contrast to their approach, the scene intrinsic reference geometry is not a noise-free and complete reference object. We therefore introduce an orientation-consistency based averaging and an adapted normal transfer approach to achieve high quality reconstructions.
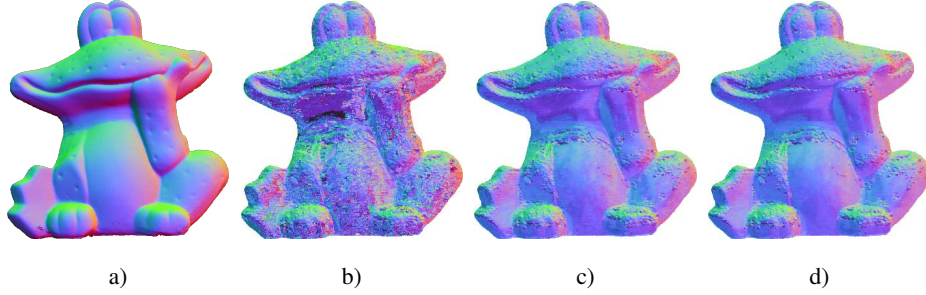
Fig. 2. Effect of best matches averaging. *a)* Ground-truth normal map, *b)* reconstructed normal map using only a single best match, *c)* average computed over 50 best matches, *d)* average computed over 100 best matches. Note that normals on the scene-intrinsic reference geometry (Figure 5b) remain unchanged.

4.1 Basic Matching

We first manually segment the target object in the images I^{PS} from the background and then project the reference geometry into I^{PS}. All pixels are classified into those covered by the scene-intrinsic reference geometry $\mathbf{Q}$ and those for which no reconstruction is available $\mathbf{P}$. Each pixel $q \in \mathbf{Q}$ is assigned a unique normal $\mathbf{n}(q)$ by projecting the reference geometry's vertices onto q. If multiple vertices are mapped to the same pixel, we choose the normal of the vertex with highest reconstruction confidence. We furthermore define the observation vector for each point in $\mathbf{P}$ and $\mathbf{Q}$ which is formed by all the color values for this particular pixel location in the image stack I^{PS}:

$$V_{p,c} = (I^{\text{PS}}_{1,p,c}, \ldots, I^{\text{PS}}_{n,p,c})^T, \quad V_{q,c} = (I^{\text{PS}}_{1,q,c}, \ldots, I^{\text{PS}}_{n,q,c})^T, \quad c \in \{R, G, B\}. \quad (2)$$

The core of geometry completion is the appropriate transfer of normals derived from the scene intrinsic reference geometry to positions where reconstruction is missing. We define the following metric for the similarity between two observation vectors that models differences of surface albedo using a per-color channel material coefficient $m_{p,c}$:
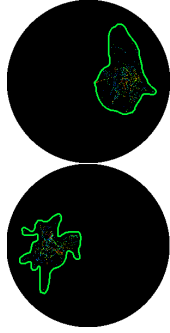
$$\Delta = \sum_{c \in \{R,G,B\}} \|m_{p,c} V_{q,c} - V_{p,c}\|_2^2 \quad (3)$$

For a given target point $p \in \mathbf{P}$, we first determine for each $q \in \mathbf{Q}$ optimal per-color channel material coefficients $m_{p,c}$:

$$m_{p,c} V_{q,c} = V_{p,c} \quad \Leftrightarrow \quad m_{p,c} V_{q,c}^T V_{q,c} = V_{q,c}^T V_{p,c} \quad \Leftrightarrow \quad m_{p,c} = \frac{V_{q,c}^T V_{p,c}}{V_{q,c}^T V_{q,c}}. \quad (4)$$

In order to find the best matching observation vector, we then select the q for which the residual error Δ in Equation 3 is minimal. We apply these steps for all points $p \in \mathbf{P}$ where reconstruction is missing.

Figure 2 (b) shows the resulting normal map for the *frog* example. Note that normals show strong artifacts in filled-in regions. This could be due to several reasons: First,

Algorithm Overview

1. Reconstruct SIRG using MVS.
2. For each pixel p in the object's mask, i.e. $\mathbf{P} \cup \mathbf{Q}$, do:
   - Compute $m_{p,c}$ for all $q \in \mathbf{Q}$.
   - Select the $s = 50$ matches $\{q_i\} \subset \mathbf{Q}$ that have minimal error

$$\sum_{c \in \{R,G,B\}} \|m_{p,c} V_{q,c} - V_{p,c}\|_2^2 \,.$$

   - Average normals for $\{q_i\}$ and transfer result to p.
3. Integrate normals.

Fig. 3. *Left:* Distribution of normal directions for 500 best matches of two target points $p, \tilde{p}$ (boundaries manually drawn for clarity). *Right:* Summary of proposed photometric stereo algorithm.

some normal directions are not represented in the reference geometry (including some individual directions but also most of the downward pointing normals on the frog's neck area). Second, the material coefficient m_p can only model differences in albedo but is unable to adapt the specularity by mixing multiple observation vectors (as in [1]). Third, even if the matching according to the orientation-consistency cue is correct, the reference geometry can still contain erroneous normal information.

4.2   Averaging Multiple Matches

If we look at a plot of the normal directions corresponding to the s best matching observation vectors for a given p (Figure 3), we notice that these are spread out over a range of directions due to the various errors in our approximation. We can, however, also observe that those normals are clustered around an average direction.

   We therefore propose to not only use the normal corresponding to the best-matching observation vector but to compute an average normal from the s best matches. This reduces the impact of wrong matches and erroneous normals and can interpolate missing normals. Note that it will not fix the case of normal directions outside the convex hull of normals observed in the scene intrinsic reference geometry but may at least assign a nearby normal direction inside the convex hull. Figure 2 shows the effect of averaging multiple matches for the frog model. Averaging the 50 best-matching normals yields a much smoother normal field. Increasing the number to 100 leads only to a small improvement. We therefore use in all cases shown in this paper $s = 50$.

4.3   Global Matching

So far, we only transferred normals from $\mathbf{Q}$ to $\mathbf{P}$. This assumes, however, that the scene intrinsic reference geometry is reconstructed with high quality which is typically not the case (see, e.g., Figure 4 showing an example of a bronze bust). Even after Laplacian smoothing, the scene-intrinsic reference geometry still contains very noisy normals.
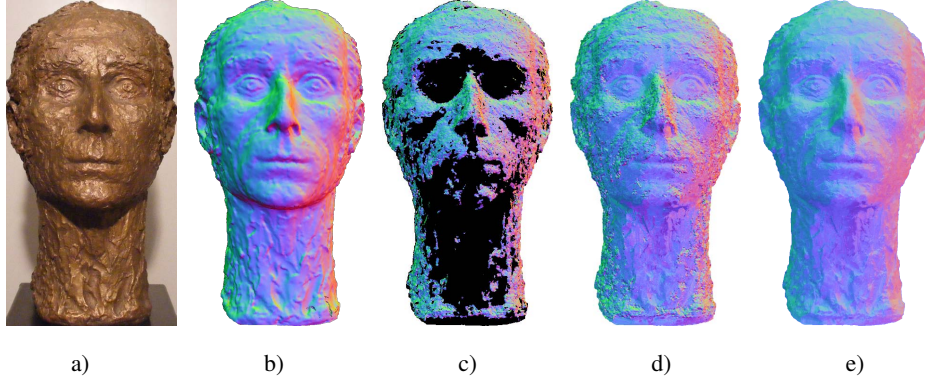
a)   b)   c)   d)   e)

Fig. 4. Effect of global matching. *a)* Example input image, *b)* ground-truth normal map, *c)* SIRG normal map, *d)* reconstructed normal map with best matches averaging only on **P**, *e)* reconstructed normal map with global matching.

We therefore apply the orientation-consistency based averaging described in Section 4.2 not only to unreconstructed regions but also to the scene-intrinsic reference geometry, thereby discarding the originally reconstructed normals. More formally, we adapt the matching in Section 4.1 to transfer normals from **Q** to both **P** *and* **Q**. In contrast to [1], Equation 3 is then minimized not only for $p \in \mathbf{P}$, but for $p \in \mathbf{P} \cup \mathbf{Q}$ with material coefficients computed for all pairs $(p, q) \in (\mathbf{P} \cup \mathbf{Q}) \times \mathbf{Q}$. This considerably improves the resulting normal map as can be seen in Figure 4 e).

5   Normal Field Integration

Several methods have been developed to integrate normal maps to recover a 3D surface (e.g. [19–23]). To constrain the possible solutions, some works propose to impose consistency with sparsely given control points from a laser scanner [24], with a visual hull [9], or with a complete depth map [25].

We follow a similar, optimization based approach as in [25] and [9]. Both operate in a perspective setting, i.e., optimize for surface points $R = (Z \cdot r_x, Z \cdot r_y, Z)$ determined by their depth Z along the ray $(r_x, r_y, 1)$. Instead of directly comparing the difference of optimized normals $\mathbf{n}_p$ to reconstructed normals $\bar{\mathbf{n}}_p$, they use the dot product between the tangent to the optimized surface and the given normal as an error metric.

Nehab et al. [25] additionally propose to introduce per pixel weights for positional and normal constraints. We use w_p as geometry weight ($w_p = \lambda$ for **Q**, $w_p = 0$ for **P**) and $u_p = v_p = 1$ as gradient weights respectively. The error function is then given as a sum over all N pixels $p \in I_1^{\mathrm{PS}}$:

$$E = \sum_p \left[u_p^2 \left\| \bar{\mathbf{n}}_p \cdot \frac{\partial Z}{\partial x} \right\|^2 + v_p^2 \left\| \bar{\mathbf{n}}_p \cdot \frac{\partial Z}{\partial y} \right\|^2 + w_p^2 \left\| \bar{Z}_p - Z_p \right\|^2 \right] \tag{5}$$

where $\bar{Z}_p$ is the depth of a reconstructed pixel. Approximating the partial derivatives with finite differences, the whole system can be written as a least squares problem with
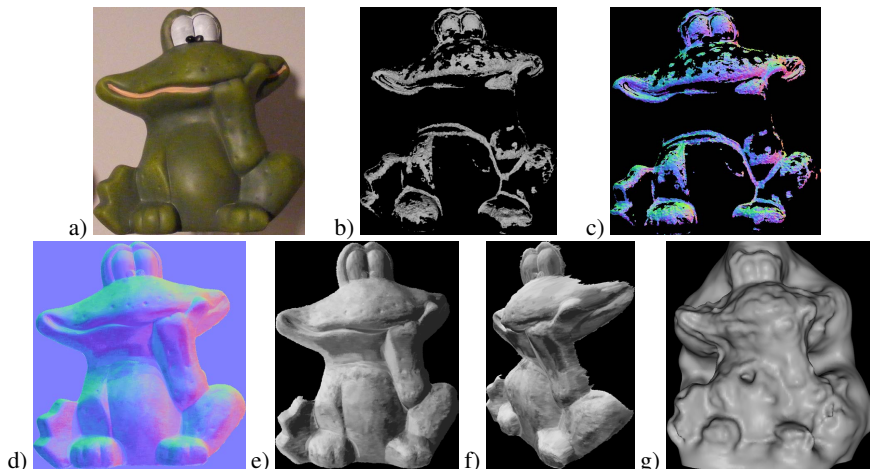
Fig. 5. *Frog* model. *a)* Example input image, *b)* partially reconstructed geometry using multi-view stereo serving as scene-intrinsic reference geometry (SIRG), *c)* SIRG normal map, *d)* reconstructed normal map using PSE with SIRG as reference object, *e)* final model after integration, *f)* final model rendered from a novel viewpoint, *g)* reconstruction with Furukawa's multi-view stereo [26] and poisson surface reconstruction [27].

a sparse $3N \times N$ matrix (see [25] for details). Because our weighting scheme does not exclude any gradients from the integration, the matrix has full rank and there exists a unique solution to the least squares problem.

6   Results

Since it is difficult to acquire ground truth data for large-scale objects, we first present a quantitative analysis on small objects that can be easily captured in a laboratory setting. We then give a qualitative evaluation for two large-scale datasets, reconstructed partially or completely from Internet images.

6.1   Lab-based Datasets

We demonstrate results on three different datasets captured under lab conditions. The *frog* is a roughly 25 cm tall clay figure with a close to diffuse surface (see Figure 5). The scene-intrinsic reference geometry covers 34 % of the foreground region $\mathbf{P} \cup \mathbf{Q}$ in the normal maps. The *bunny* is a plastic figurine with shiny coating (about 20 cm tall, 47 % coverage, see Figure 6). The bronze *bust* (40 cm tall, 53 % coverage, see Figure 4 a)) exhibits complex surface structure and a difficult BRDF.

The datasets were all acquired using a 7 M pixel consumer camera. We captured 15-20 I^{PS} images from a fixed camera position while manually moving a simple light bulb around the object. We additionally captured ≈ 50 I^{MVS} images from various positions

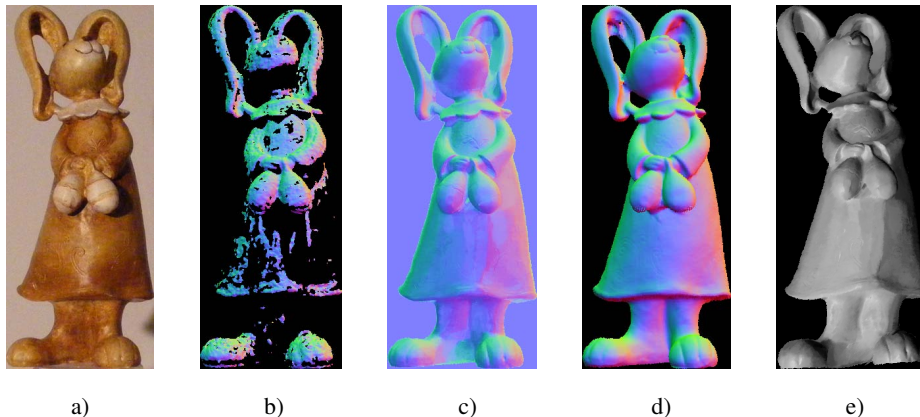<div align="center">a)   b)   c)   d)   e)</div>

Fig. 6. *Bunny* model. *a)* Example input image, *b)* normal map of partially reconstructed geometry using multi-view stereo serving as scene-intrinsic reference geometry (SIRG), *c)* reconstructed normal map, *d)* ground truth, *e)* final model rendered from a novel viewpoint.

facing the front side of the objects. Neither camera nor light source were calibrated. In order to evaluate our reconstructions quantitatively, we scanned the objects using a structured light scanner. From the merged and cleaned point model, we created ground-truth normal maps for comparison. Note that these ground-truth normal maps show holes in areas where scanning was difficult due to self-occlusion, e.g., the bunny's ear region in Figure 6.

Evaluation Figures 5, 4, and 6 demonstrate clearly that our approach is able to reconstruct high quality normal maps without requiring special reference objects in the scene. Even small details such as the flowers on the bunny's dress are reconstructed. Normals outside the convex hull of captured normal directions such as the chin area in Figure 4 or areas with self shadowing around the bust's nose are reconstructed plausibly without introducing strong artifacts. The final integrated models are of high quality and avoid large scale distortions due to the inclusion of reference geometry in the integration routine.

Figure 7 shows for all lab datasets histograms over the deviation of reconstructed normals compared to normals computed from the scanned ground truth model. The graphs clearly show that normals obtained from multi-view stereo techniques are improved by our proposed normal transfer.

To demonstrate that our technique works also with more general lighting, we captured two additional datasets for the *bust* where we used a studio light with and without diffuser for illumination. The close-ups in Figure 8 demonstrate that the recovered normals change only marginally. Note that this is a key requirement for applying the technique to outdoor scenes as shown in the next section.

It does not matter to our approach in which manner the scene intrinsic reference geometry is obtained. We therefore additionally applied the multi-view stereo algorithm of Furukawa and Ponce [26] to the *bunny* and *frog* datasets. The resulting point
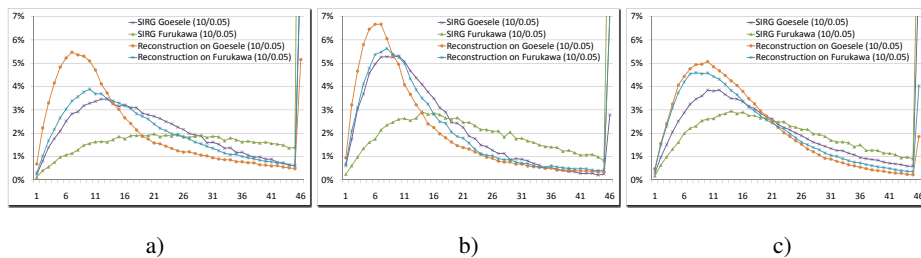
a)

b)

c)

Fig. 7. Normal deviation in degrees against ground truth for *a) frog*, *b) bunny*, and *c) bust*. The purple (green) line shows the deviation of the SIRG obtained by multi-view stereo of Goesele et al. (Furukawa et al.) after Laplacian smoothing. The orange (blue) line demonstrates the result of our complete pipeline on Goesele et al. (Furukawa et al.) multi-view stereo input data. Data point 46 represents deviations greater than $45°$. Values out of scale are: a) SIRG G. 9.4%, SIRG F. 29.5%, Rec. F. 9.8%; b) SIRG F. 15.9%, Rec. G. 7.1%, Rec. F. 7.2%; c) SIRG G. 8.9%, SIRG F. 14.2%.
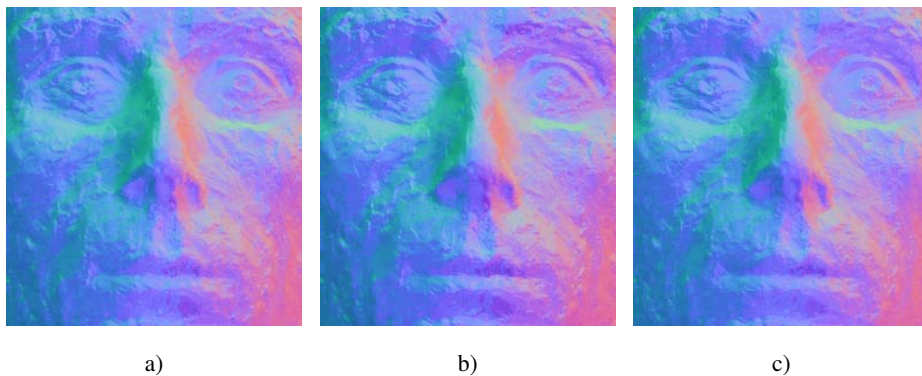

a)

b)

c)

Fig. 8. Reconstruction of *bust* for light situations *a)* diffuse spot, *b)* bright spot, and *c)* point light.

set was then used as input to our adapted photometric stereo by example technique. Figure 7 shows the deviation of the results from groundtruth. Like for the input from Goesele et al., we observe a significant improvement (blue lines) of the normals (green lines) through our matching scheme. Furthermore, the figure shows that input normals from Furukawa's method (green lines) are farther away from the groundtruth than those reconstructed by VRIP and Goesele's method (purple lines). Obviously, this leads to the differences in the resulting normal map's quality (comparing the orange and blue lines). We also attempted to reconstruct a triangle mesh from the point cloud created by Furukawa's method using Poisson surface reconstruction [27] but despite several trials with different parameter settings the Poisson reconstruction did not yield satisfying results (see Figure 5 g) for an example).

<div style="text-align:center">a)   b)</div>

Fig. 9. *Cathedral* dataset. *Left:* Images captured by a static webcam used as I^{PS}. *Right:* Images downloaded from a community photo collection site used as I^{MVS}.

<div>a)   b)   c)</div>
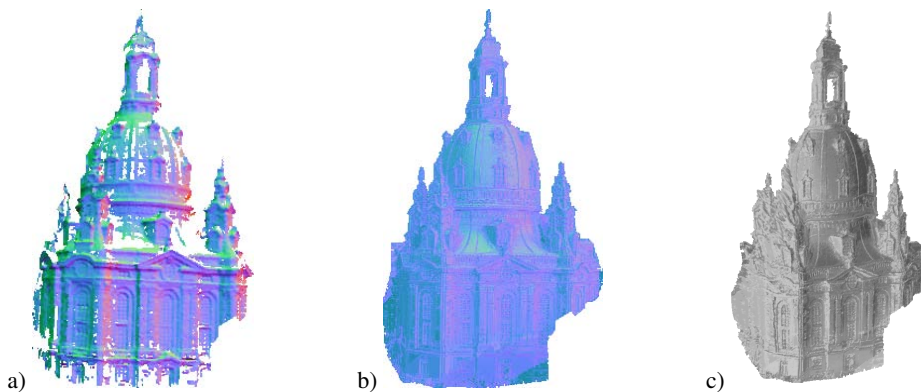
Fig. 10. *Cathedral* model. *a)* Normal map of partially reconstructed geometry using multi-view stereo serving as scene-intrinsic reference geometry (SIRG), *b)* reconstructed normal map, *c)* final model rendered from novel viewpoint.

6.2 Outdoor Datasets

We discuss the performance on outdoor scenes using two large buildings (about 60 m and 90 m tall) with non-planar surfaces and interesting details. For each dataset, we retrieved an image of a public webcam every 20 min over the course of 3 months. The webcam images have VGA resolution.

We manually selected 11 suitable I^{PS} images for the *cathedral* dataset (all taken between 10 am and 5 pm) and 36 images for the *tower* dataset (taken between 9 am an 7 pm) on different days, see Figures 9 a), 1 a). For the *cathedral*, we furthermore downloaded 2000 I^{MVS} images from the community photo platform Flickr (see Figure 9 b)). The multi-view stereo step automatically selected a suitable subset of those for reconstruction of the scene intrinsic reference geometry achieving a completeness of 84 %. The SIRG for the *tower* was reconstructed from 324 images taken by a student with a consumer camera (see Figure 1 b)) and covers 58 % of the foreground.

Evaluation Figure 10 demonstrates the results for the *cathedral* model. The fairly complete reference geometry is a good basis for reconstruction. The global matching softens the extremes in the normal map but strongly increases the available detail. Artifacts can be seen due to cast shadows on the object that violate the assumption of distant illumination and are not modeled by our approach (e.g., lower right corner of Figure 10 c)).

As a cylindrical object, the *tower* is well-suited for reconstruction. Parameterizing its surface by height $h \in [0, H]$ and angle φ, a normal at (h, φ) can be reconstructed quite accurately if some normal on the line $([0, H], \varphi)$ is contained in the reference geometry. This works so well that the bottom of the tower can be recovered up to fine details like individual stones (see Figure 1 d)). Even if the roof has a different albedo and only sparse coverage of normal directions pointing to the right, we are able to reconstruct it quite convincingly.

7   Discussion

Being able to reconstruct sufficient geometry for the SIRG is a key requirement of our algorithm. However for some scenes, it will most likely be impossible to reconstruct sufficient geometry with *any* multi-view stereo algorithm. Such scenes need to be treated differently. We argue, however, that this is a rare case and that the chosen multi-view stereo algorithm [17] (or another MVS approach) will for most scenes be able to reconstruct at least some geometry. We demonstrated that this geometry can be used as reference geometry, bootstrapping the photometric stereo by example approach. This yields the clear benefit that neither lighting nor scene reflectance need to be known or even controlled.

Another critical point is the reasoning in Section 4.2 why averaging multiple matches works. It is, e.g., clear that averaging multiple normal directions will not handle a mismatch in reflectance between a point and the scene intrinsic reference geometry as well as mixing the contributions of two reference objects in [1] would do. We found, however, that it is a procedure based on the available information that in practice yields surprisingly good results.

8   Conclusion and Future Work

Reconstructing accurate normals for large-scale objects with photometric stereo methods is a non-trivial task. As we cannot put whole buildings in a laboratory, many traditional photometric stereo methods cannot be easily applied. In this paper we presented a combination of multi-view stereo and photometric stereo that is able to cope with outdoor imagery and has minimal capturing requirements. Like standard photometric stereo by example, it neither requires known lighting or reflectance nor does it reconstruct either of them explicitly. By introducing the scene-intrinsic reference geometry, we are able to extend the applicability of photometric stereo by example to scenes for which it is undesired or even impossible to include reference objects.

The reference geometry can be seen as a set of noisy samples of the function f from observation vectors to normals. The current best matches averaging does not consider

how close the individual matches actually are to the candidate and it does not model the noise in the input data. This could be remedied by interpolation methods like Kriging that estimates unknown values based on known values at nearby points. However, better understanding of the space of observation vectors and their distribution is needed and we will further explore this field. A starting point is already provided by Sato et al. [28] who investigate similarity measures for observation vectors and apply a dimension reduction technique to the space of observation vectors.

In the future, we would furthermore like to improve robustness against cast shadows and local influences. Finally, the reconstructed 3D geometry might benefit from better integration, e.g., by avoiding to integrate over depth discontinuities similar to Vlasic et al. [9] or Agrawal et al. [29].

Acknowledgments

References

1. Hertzmann, A., Seitz, S.M.: Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. PAMI **27** (2005) 1254–1264
2. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Optical Engineering **19** (1980) 139–144
3. Horn, B.: Robot Vision. MIT Press (1986)
4. Basri, R., Jacobs, D.: Photometric stereo with general, unknown lighting. In: Proc. ICCV. (2001)
5. Ramamoorthi, R., Hanrahan, P.: A signal-processing framework for inverse rendering. In: Proc. SIGGRAPH. (2001) 117–128
6. Shen, L., Tan, P.: Photometric stereo and weather estimation using internet images. In: Proc. CVPR. (2009)
7. Joshi, N., Kriegman, D.: Shape from varying illumination and viewpoint. In: Proc. CVPR. (2007)
8. Hernandez, C., Vogiatzis, G., Cipolla, R.: Multi-view photometric stereo. PAMI (2008)
9. Vlasic, D., Peers, P., Baran, I., Debevec, P., Popovic, J., Rusinkiewicz, S., Matusik, W.: Dynamic shape capture using multi-view photometric stereo. ACM Trans. Graph. **28** (2009)
10. Higo, T., Matsushita, Y., Joshi, N., Ikeuchi, K.: A hand-held photometric stereo camera for 3-d modeling. In: Proc. ICCV. (2009)
11. Goldman, D.B., Curless, B., Hertzmann, A., Seitz, S.M.: Shape and spatially varying BRDFs from photometric stereo. In: Proc. ICCV. (2005) 341–348
12. Aldrin, N., Zickler, T., Kriegman, D.: Photometric stereo with non-parametric spatially-varying reflectance. In: Proc. CVPR. (2008)
13. Silver, W.M.: Determining shape and reflectance using multiple images. Master's thesis, Massachusetts Institute of Technology (1980)
14. Koppal, S.J., Narasimhan, S.G.: Clustering appearance for scene analysis. In: Proc. CVPR. (2006) 1323–1330
15. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3D. ACM Trans. Graph. **25** (2006) 835–846

16. Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proc. CVPR. (2006) 519–526 http://vision.middlebury.edu/mview/.
17. Goesele, M., Snavely, N., Curless, B., Hoppe, H., Seitz, S.M.: Multi-view stereo for community photo collections. In: Proc. ICCV. (2007)
18. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: Proc. SIGGRAPH. (1996) 303–312
19. Wu, Z., Li, L.: A line-integration based method for depth recovery from surface normals. Computer Vision, Graphics, and Image Processing **43** (1988) 53–66
20. Klette, R., Schluens, K.: Height data from gradient fields. In: Proc. Machine Vision Applications, Architectures, and Systems Integration V. (1996) 204–215
21. Smith, W.A.P., Hancock, E.R.: Statistical methods for surface integration. In: Proc. IMA Conference on the Mathematics of Surfaces. (2007) 427–441
22. Ho, J., Lim, J., Yang, M.H., Kriegman, D.: Integrating surface normal vectors using fast marching method. In: Proc. ECCV. (2006) 239–250
23. Durou, J.D., Aujol, J.F., Courteille, F.: Integrating the normal field of a surface in the presence of discontinuities. In: Proc. EMM-CVPR. (2009) 261–273
24. Horovitz, I., Kiryati, N.: Depth from gradient fields and control points: Bias correction in photometric stereo. Image and Vision Computing **22** (2004) 681–694
25. Nehab, D., Rusinkiewicz, S., Ramamoorthi, R.: Efficiently combining positions and normals for precise 3d geometry. ACM Trans. Graph. **24** (2005) 536–543
26. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. PAMI **32** (2010)
27. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Proc. SGP. (2006) 61–70
28. Sato, I., Okabe, T., Yu, Q., Sato, Y.: Shape reconstruction based on similarity in radiance changes under varying illumination. In: Proc. ICCV. (2007) 1–8
29. Agrawal, A., Raskar, R., Chellappa, R.: What is the range of surface reconstructions from a gradient field? In: Proc. ECCV. (2006)