Robust Outlier Removal from Point Clouds Acquired with Structured Light

J. Köhler[1], T. Nöll[1], G. Reis[1] and D. Stricker[1]

[1]German Research Center for Artificial Intelligence, Trippstadter Str. 122, D-67663 Kaiserslautern, Germany

Abstract

We present a method for automatic, robust identification of outliers in point clouds acquired with structured light. In contrast to most state of the art methods, we consider available 2D information instead of operating on the points only. Our method performs robust even for complex, glossy surfaces, where illumination artifacts introduce ghost geometry or outliers that seamlessly blend into the correct surface. The key idea of our algorithm is to only use cameras for reconstruction and the projector for a separate consistency check. We use encoding by phase shifting to obtain per pixel correspondences for the cameras and subpixel accuracy for the projector. The method requires at least 2 cameras, a projector and a precise calibration of all devices.

Categories and Subject Descriptors (according to ACM CCS): I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Scanning

1. Introduction

The usability of a reconstructed point cloud is significantly decreased the more outliers it contains. Especially when complex, shiny surfaces are reconstructed, global illumination artifacts and specular lobes can cause ghost geometry and outliers that seamlessly blend into the correct surface or are closely distributed around it (Figure 1, top). In order to identify those outliers, existing state of the art postprocessing methods often operate on the 3D geometry only [SBS05, WPH*04, YCC09]. However, the results are often not satisfying: Outliers close to correct points might be treated as noise and smoothed into the final surface, while ghost geometry is itself consistent and cannot be distinguished from the correct parts.

Instead of operating on the 3D points only, we additionally use 2D information that is provided by the calibration of the scanning devices for filtering. Image based reconstruction methods usually establish dense 2D correspondences among the input images. A point cloud is then reconstructed by computing 3D points that yield small reprojection errors when backprojected to these correspondences [HZ04]. The presence of the aforementioned outliers however shows, that a substantial amount of points yields a small reprojection error but is nevertheless wrong. We address this problem by using a structured light scanning setup, where only cameras are used for computing the 3D geometry. The projector
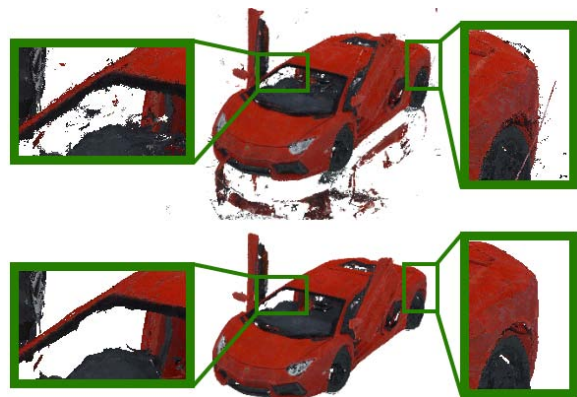

Figure 1: *A glossy toy car scanned with structured light while standing on a glass plate. Top row: Unfiltered data, bottom row: Geometry filtered with our method.*

does not participate in the triangulation, but serves as an additional control instance. A vertex reconstructed within the bounds of the camera reprojection error is verified by checking its reprojection error with the projector. On the one hand, this circumvents that the nonlinear gamma transfer function between a camera and a projector introduces errors to the reconstruction. On the other hand, the additional information

Figure 2: *Left to right: The projected Φ_x/Φ_y (red/blue), fringe encoding, ϕ_i/ϕ_j with exemplary correspondences for two cameras C_i, C_j.*

Figure 3: *Triangulation with a projector with uncalibrated gamma curve. Green: Ground truth, red: Oscillating reconstruction.*

constituted by the projector is not disregarded, but serves as a powerful outlier filter. This makes our algorithm robust also in regions close to the correct surface. Outliers are identified with excellent precision, while the correct parts of the reconstructed point clouds are left almost untouched. The point clouds can thus be used for further processing immediately.

2. Geometry Acquisition

The significance of the reprojection error that a vertex scores when it is projected onto the image plane of a projector strongly depends on the granularity of the projected patterns. We thus use phase shifted structured light, which yields correspondences at the highest possible resolution: Per pixel for the cameras and with subpixel accuracy for the projector. Two phase functions $\Phi_x(x,y) = (x,x)^T$ and $\Phi_y(x,y) = (y,y)^T$ $x,y \in [0..1]$ are transmitted from the projector to all cameras by encoding them into sinusoidal fringe patterns that are shifted n times. In the following, $\phi_i = (\phi_{ix}, \phi_{iy})$ denotes the recovered phase as seen by camera C_i, $\phi_i(p) \in [0..1] \times [0..1]$ for a camera pixel p (Figure 2). Several methods exist for recovering ϕ_i from n phase shifts, we use the least squares approach discussed in [GZC07]. This method yields ϕ_i wrapped to the respective intervals. We unwrap it by applying the temporal phase unwrapping approach proposed in [WNB10].

Correspondences between a pixel p_i of camera C_i and a pixel p_j of camera C_j, $i \neq j$ can be deduced from ϕ_i, ϕ_j with the following criterion:

$$\left|\left|\phi_i(p_i) - \phi_j(p_j)\right|\right| < \varepsilon_\Phi \qquad (1)$$

We reconstruct those parts of the surface that are lit by the projector and seen by multiple cameras using a method similar to [HLZ10]. For each pixel p_i of each camera C_i that has a valid phase $\phi_i(p_i)$, we compute a vertex V on the ray r_{pi} through p_i. V is initialized by triangulating a pixel correspondence p_j to another camera C_j that minimizes Equation 1. The depth of V on r_{pi} is then optimized iteratively among all m cameras where the projection of V fulfills Equation 1, using the Levenberg Marquardt algorithm. As last step, the phase values $\phi_j(p_j)$ of all m observing cameras are attached to V.

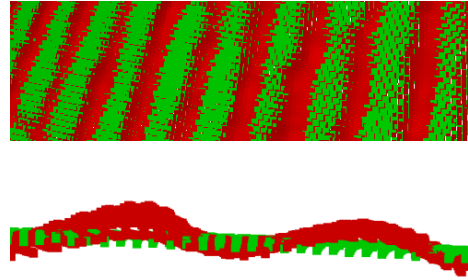Note, that in contrast to other methods (e.g. [ZH04]), we do not use the projector for triangulation. The major reason for this is the nonlinear gamma transfer function between the projector and a camera, that would be propagated into the geometry within the bounds of the reprojection error. As a result, the fringe oscillations are visible in the reconstruction (Figure 3). A possible solution to this problem would be the calibration of the gamma transfer function [WNB10]. This works, however, only for a single camera-projector combination, because it changes the projected pattern with respect to a single camera. Moreover, we found that a gamma calibration can reduce the spectrum of values that are projected, depending on the response of the projector's actual gamma function and the sensitivity of the camera used.

3. Filtering

A point cloud reconstructed by the algorithm described in Section 2 is not affected by the nonlinear gamma transfer function of the projector, but can still contain many outliers. Especially when reconstructing glossy surfaces, view dependent illumination artifacts like specular lobes and flares cause phase errors for a camera in regions where other cameras observe the correct values. Those errors can emerge continuously, due to the fading of a lobe. Correspondences among the cameras in such a region are wrong, but often within the bound of Equation 1. Reconstruced points are not necessarily noisy, but can form a consistent, displaced surface patch, that seamlessly blends into the correct surface in the worst case (Figure 4). Another typical, unwanted artifact is ghost geometry caused by indirect illumination that is consistently observed by all cameras. When the projector is used as a filtering instance, such outliers can be identified with excellent precision, while the correct parts of the geometry are left almost untouched.

The formation of outliers is illustrated in Figure 5. The left image shows, how phase values are distributed among all devices when no error occurs. Light projected onto a surface is consistently reflected to the cameras. In the right image, C_1 captures a wrong phase value t. Consequently, the
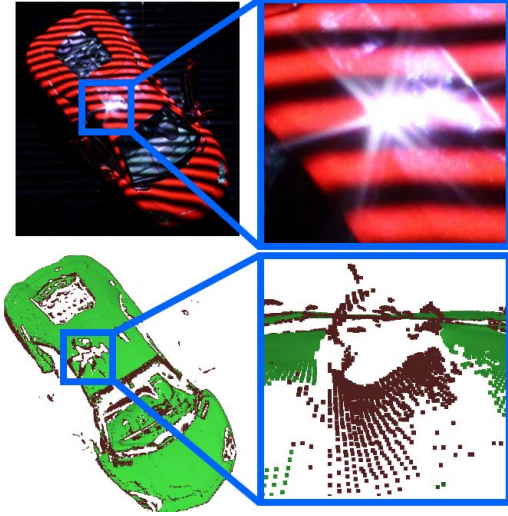
Figure 4: *Typical errors introduced by a flare. Valid geometry is green, invalid geometry is red in the bottom row.*
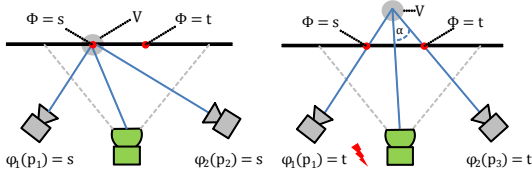

Figure 5: *Left: No error, right: C_1 observes a wrong phase value t instead of s.*

pixel correspondence $p_1 \leftrightarrow p_3$ with C_2, that was established using the phase values is wrong. When p_3 lies close to the epipolar line induced by p_1 however, the reprojection error is low enough for a successful triangulation of a vertex V. It is obvious from Figure 5 (right), that V will still score a bad reprojection error with the projector, because its ray to V does not pass through the correct location of t on the surface.

As described in Section 2, V is associated with m phase values ϕ_i, when it is observed by m cameras. These values vary within the bound of ε_Φ (Equation 1). For each ϕ_i, the pixel of the projector that should have illuminated V is computed as

$$p_{\phi_i} = (\phi_{ix} \cdot res_x, \phi_{iy} \cdot res_y)^T \qquad (2)$$

where $res_{x/y}$ is the resolution of the projector. Note that p_{ϕ_i} is computed with subpixel accuracy because the projected, continuous (Φ_x, Φ_y) is known. All m phase values thus yield m slightly different pixels for the projector.

We consider V as invalid, when its projection into the image plane of the projector scores a reprojection error larger
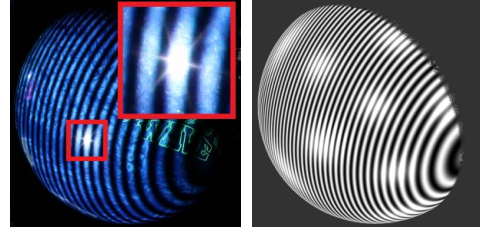

Figure 6: *left: Fringes projected onto the bowling ball, the polished surface exhibits flares and small, coarse grained illumination artifacts. Right: Synthetic fringe image with specular lobes.*

than ε_p with one of the m reference pixels:

$$\left|\left|\mathbb{P}(V) - p_{\phi_i}\right|\right| > \varepsilon_p \qquad (3)$$

where $\mathbb{P}(V)$ denotes the projection of V into the image plane of the projector. The reliability of this filtering criterion with respect to a single camera depends on the included angle α_i between the projector ray and the ray of those cameras C_i, that observe the correct phase value (Figure 5). For a small α_i, V will be projected closer to the correct projector pixel, even if the vertex is wrong. This is important, when a small amount of cameras (i.e. 2 or 3) is used, which have different angles towards the projector. In this case ε_p should be weighted linearly with α_i, that it gets smaller when α_i decreases.

$$\varepsilon_p = \varepsilon'_p \cdot (1 - \frac{2 \cdot \left|cos^{-1}(r_{C_i} \cdot r_p) - \frac{\pi}{2}\right|}{\pi}) \qquad (4)$$

where r_{C_i} and r_p are the normalized camera/projector rays through V and ε'_p is a fixed reprojection threshold depending on the projector resolution.

4. Results

We test our method by applying it to both real and synthetic data. We use an acquisition setup with 7 cameras (Canon® EOS 500D) and 1 projector (Sanyo® Z4000) and project the phase encoded by fringe images with 3 and 10 phase shifts. The real datasets are scans of a bowling ball. It was burnished with car polish to increase the specular reflection component, which ensures a certain amount of errors. The synthetic data sets are scans of a perfect sphere that were simulated by ray tracing. We model global illumination artifacts in the scans by applying the Phong reflection model [Pho75] with 12 additional, static light sources (Figure 6).

We fit a sphere to both the filtered and the original, outlier afflicted datasets using RANSAC [FB81] for outlier rejection. The sphere parameters derived from the best RANSAC consensus set are optimized iteratively by applying the Levenberg Marquardt algorithm. For each fit we then compute

	radius	mean diff	max diff	σ	#points	%removed	ε_p
Synthetic sphere, 3 phase shifts							
unfiltered	180	0.01717	99.84	0.1534	19113296	8.2%	0.5
filtered	180	0.009356	0.239	0.01625	17542392		
Synthetic sphere, 10 phase shifts							
unfiltered	180	0.009501	59.63	0.05582	19457187	1.9%	0.5
filtered	180	0.009238	0.06865	0.01557	19068114		
Bowling ball, 3 phase shifts							
unfiltered	99.05	0.09869	7.54	0.166	5605980	18%	0.7
filtered	99.01	0.08985	0.7928	0.1104	4591426		
Bowling ball, 10 phase shifts							
unfiltered	99.01	0.08393	4.692	0.1574	6040120	9.5%	0.7
filtered	99.02	0.06634	0.6649	0.1156	5462529		

Table 1: *Experimental results for all datasets. Mean and max diff abbreviate the mean and maximum difference to the radius.*

the average and maximum deviation from the radius and the standard deviation σ. The results are listed in Table 1.

Three important conclusions can be drawn from the listed data. First, the amount of filtered vertices is lower for the datasets with 10 phase shifts. The least squares phase estimation tends to yield better phase values ϕ_i, when more shifts are used. This is also obvious when comparing the standard deviation of the unfiltered 3 and 10 shift datasets with each other: σ is distinctly lower, when 10 shifts are used. Second, the filtered datasets yield a significantly lower σ than the unfiltered datasets. This indicates that the point cloud filtered with our method is less noisy and has a better quality than the unfiltered point cloud. Third, the standard deviations for the filtered data sets are very similar to each other for different amounts of phase shifts. This holds for both real and synthetic datasets and demonstrates the sustainable quality of the filtering. Even with only 3 phase shifts, the filtered sphere has similar characteristics than the one acquired with 10 shifts.

Two types of values are remarkably high: The peaks of the radius deviation in the unfiltered, synthetic datasets result from single, large outliers. The filtering rate for the bowling ball scanned with 3 shifts is large, because its reconstructed border is noisy. The ball's surface is very challenging for a diffuse structured light scanner because it has, in addition to its glossiness, a metallic like finish that causes coarse grained noise (Figure 6, left) in the fringe images.

5. Conclusion

We have shown that the projector of a structured light scanning setup can be used as a powerful instance for filtering outliers. This is more appropriate than integrating it in the triangulation process, because then the nonlinear gamma function is not propagated into the reconstructed points. The method removes almost all outliers, but leaves the correct parts of the geometry untouched.

References

[FB81] FISCHLER M. A., BOLLES R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24* (June 1981), 381–395. 3

[GZC07] GUO H., ZHAO Z., CHEN M.: Efficient iterative algorithm for phase-shifting interferometry. *Optics and Lasers in Engineering 45*, 2 (2007), 281 – 292. 2

[HLZ10] HOLROYD M., LAWRENCE J., ZICKLER T.: A coaxial optical scanner for synchronous acquisition of 3d geometry and surface reflectance. In *ACM SIGGRAPH 2010 papers* (New York, NY, USA, 2010), SIGGRAPH '10, ACM, pp. 99:1–99:12. 2

[HZ04] HARTLEY R. I., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press, ISBN: 0521540518, 2004. 1

[Pho75] PHONG B. T.: Illumination for computer generated pictures. *Commun. ACM 18* (June 1975), 311–317. 3

[SBS05] SCHALL O., BELYAEV A., SEIDEL H.-P.: Robust filtering of noisy scattered point data. In *IEEE/Eurographics Symposium on Point-Based Graphics* (Stony Brook, New York, USA, 2005), Pauly M., Zwicker M., (Eds.), Eurographics Association, pp. 71–77. 1

[WNB10] WANG Z., NGUYEN D. A., BARNES J. C.: Some practical considerations in fringe projection profilometry. *Optics and Lasers in Engineering 48*, 2 (2010), 218 – 225. 2

[WPH*04] WEYRICH T., PAULY M., HEINZLE S., SCANDELLA S., GROSS M.: Post-Processing Of Scanned 3D Surface Data. In *IEEE/Eurographics Symposium on Point-Based Graphics* (2004), pp. 85–94. 1

[YCC09] YANG R.-Q., CHENG S., CHEN Y.-Z.: A two-step outlier removal procedure for surface reconstruction. *Journal of Shanghai Jiaotong University (Science) 14* (2009), 266–272. 10.1007/s12204-009-0266-1. 1

[ZH04] ZHANG S., HUANG P.: High-resolution, real-time 3d shape acquisition. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on* (june 2004), p. 28. 2