Shading Constraint Improves Accuracy of Time-of-Flight Measurements

Martin Böhme, Martin Haker, Thomas Martinetz, and Erhardt Barth
Institute for Neuro- and Bioinformatics, University of Lübeck
Ratzeburger Allee 160, 23538 Lübeck, Germany
http://www.inb.uni-luebeck.de

Abstract

We describe a technique for improving the accuracy of range maps measured by time-of-flight (TOF) cameras. The technique is based on the observation that the range map and intensity image measured by a TOF camera are not independent but are linked by the shading constraint*: If the reflectance properties of the surface are known, a certain range map implies a corresponding intensity image.*

We impose the shading constraint using a probabilistic model of image formation and find a maximum a posteriori estimate for the true range map. We present results on both synthetic and real TOF camera images that demonstrate the robust shape estimates achieved by the algorithm.

1. Introduction

The time-of-flight (TOF) camera provides a range map that is perfectly registered with an intensity image (often called an *amplitude* image in TOF nomenclature), making it an attractive sensor for a wide range of applications.

In this paper, we present a technique for improving the accuracy of the TOF camera's range measurements, based on the insight that the range and intensity measurements are not independent, but are linked by the *shading constraint*: Assuming that the reflectance properties of the object surface are known, we can deduce the intensity image that should be observed. In practice, a general reflectance model (such as Lambertian reflectance) will provide an acceptable approximation to the properties of a wide range of objects.

In theory, the shading constraint can be used to reconstruct the range map from an intensity image alone; this idea has been exploited in a wide range of *shape from shading* (SfS) algorithms (see [4] for a survey). A principal limitation of these algorithms, however, is that they cannot determine whether intensity changes are caused by the object's shape or by changes in the object's reflectivity (or *albedo*). Because of this, the object is usually assumed to have constant albedo; this limits the applicability of SfS methods.

The range map measured by the TOF camera provides valuable additional information: By placing a strong constraint on the shape of the object, it allows the shading constraint to be applied more effectively and robustly.

In previous work on fusing range and intensity data, a number of authors exploit the fact that an edge in the intensity data often co-occurs with an edge in the range data. Nadabar and Jain [8] use a Markov random field (MRF) to identify different types of edges. Diebel and Thrun [2] use edge strengths estimated on a high-resolution colour image to increase the resolution of a low-resolution depth map.

A number of researchers have also integrated the shading constraint with other range information. Most of this work focuses on the integration of SfS with stereo; for example, see the work of Thompson [12], Fua and Leclerc [5], or Samaras et al. [11]. However, these approaches usually make explicit use of stereo disparity and, thus, do not generalize to range maps computed in other ways.

Some approaches combine shading with a range map obtained by arbitrary means; stereo may be used, but it is not essential to the formulation of the algorithm. Leclerc and Bobick [6] use a stereo range map to initialize an iterative SfS method. Cryer et al. [1] use a heuristic that combines low-frequency components from the stereo range map with high-frequency components from the SfS range map. Mostafa et al. [7] use a neural network to interpolate the difference between the SfS result and a more coarsely sampled range map from a range sensor; the SfS result is corrected using this error estimate. These approaches allow arbitrary range maps to be used, but they are all somewhat ad-hoc.

Our approach to improving the accuracy of the range map using the shading constraint is based on a probabilistic model of the image formation process. We obtain a maximum a posteriori estimate for the range map using a numerical minimization technique. The approach has a solid theoretical foundation and incorporates the sensor-based range information and the shading constraint in a single model; for details, see Section 2. The method delivers robust estimation results on both synthetic and natural images, as we show in Section 3.

2. Method

2.1. Probabilistic Image Formation Model

We seek to find the range map $\mathbf{R}$ that maximizes the posterior probability

$$p(\mathbf{X}^{\mathrm{R}}, \mathbf{X}^{\mathrm{I}} | \mathbf{R}, \mathbf{A}) \, p(\mathbf{R}) \, p(\mathbf{A}). \tag{1}$$

$p(\mathbf{X}^{\mathrm{R}}, \mathbf{X}^{\mathrm{I}} | \mathbf{R}, \mathbf{A})$ is the probability of observing a range map $\mathbf{X}^{\mathrm{R}}$ and an intensity image $\mathbf{X}^{\mathrm{I}}$ given that the true range map describing the shape of the imaged object is $\mathbf{R}$ and that the parameters of the reflectance model are $\mathbf{A}$. (Typically, $\mathbf{A}$ is the albedo of the object – we will discuss this in more detail below.) $p(\mathbf{R})$ is a prior on the range map, $p(\mathbf{A})$ is a prior on the reflectance model parameters.

The conditional probability $p(\mathbf{X}^{\mathrm{R}}, \mathbf{X}^{\mathrm{I}} | \mathbf{R}, \mathbf{A})$ is based on the following model of image formation: The observed range map $\mathbf{X}^{\mathrm{R}}$ is simply the true range map $\mathbf{R}$ with additive Gaussian noise, i.e.

$$p(\mathbf{X}^{\mathrm{R}} | \mathbf{R}) = \mathcal{N}(\mathbf{X}^{\mathrm{R}} - \mathbf{R} | \mu = 0, \sigma_{\mathrm{R}}). \tag{2}$$

The shading constraint postulates that a given range map $\mathbf{R}$ is associated with an intensity image $\mathbf{I}(\mathbf{R}, \mathbf{A})$, where the function expressed by $\mathbf{I}$ depends on the reflectance model. We generally use the Lambertian reflectance model, see Section 2.2; in this case, $\mathbf{A}$ is the albedo of the object, which may vary from pixel to pixel. Again, we assume that the intensity image is corrupted by additive Gaussian noise, i.e.

$$p(\mathbf{X}^{\mathrm{I}} | \mathbf{R}, \mathbf{A}) = \mathcal{N}(\mathbf{X}^{\mathrm{I}} - \mathbf{I}(\mathbf{R}, \mathbf{A}) | \mu = 0, \sigma_{\mathrm{I}}). \tag{3}$$

For the range map prior $p(\mathbf{R})$, we use the shape prior introduced by Diebel et al. [3], which favours surfaces with smoothly changing surface normals. We tessellate the range map into triangles and compute the surface normal $\mathbf{n}_j$ for each triangle. The shape prior is then given by the energy function

$$E^{\mathrm{R}}(\mathbf{R}) = w_{\mathrm{R}} \sum_{\substack{\text{triangles } j,k \\ \text{adjacent}}} \|\mathbf{n}_j - \mathbf{n}_k\|_2, \tag{4}$$

which implies the distribution $p(\mathbf{R}) = \frac{1}{Z} \exp(-E^{\mathrm{R}}(\mathbf{R}))$, where Z is a normalization constant. w_{R} is a constant that controls the dispersion of the distribution.

We now turn to the prior $p(\mathbf{A})$ for the parameters $\mathbf{A}$ of the reflectance model. In the Lambertian reflectance model, these are the albedo values at each pixel location. We will investigate several alternatives for the prior $p(\mathbf{A})$: (i) A single albedo value, specified beforehand, is used for all pixels. (ii) The same global albedo is used for all pixels, but its value is allowed to vary; we assume a uniform distribution for this global albedo. (iii) Each pixel location may have a

different albedo, and the prior $p(\mathbf{A})$ favours smooth albedo changes. In this latter case, we use an energy function

$$E^{\mathrm{A}}(\mathbf{A}) = w_{\mathrm{A}} \sum_{\substack{\text{pixels } j,k \\ \text{adjacent}}} |a_j - a_k|, \tag{5}$$

which implies the prior $p(\mathbf{A}) = \frac{1}{Z} \exp(-E^{\mathrm{A}}(\mathbf{A}))$, in analogy to the shape prior defined above.

As usual, we take the negative logarithm of the likelihood function and eliminate constant additive terms to obtain an energy function

$$\begin{aligned} E(\mathbf{R}, \mathbf{A}) = & \sum_j \frac{(X_j^{\mathrm{R}} - R_j)^2}{2\,\sigma_{\mathrm{R}}^2} + \\ & \sum_j \frac{(X_j^{\mathrm{I}} - I_j(\mathbf{R}, \mathbf{A}))^2}{2\,\sigma_{\mathrm{I}}^2} + \\ & E^{\mathrm{R}}(\mathbf{R}) + E^{\mathrm{A}}(\mathbf{A}), \end{aligned} \tag{6}$$

where the index j runs over all pixels.

We find the maximum likelihood estimate for the range map by minimizing $E(\mathbf{R}, \mathbf{A})$ using the Polak-Ribière variant of the nonlinear conjugate gradient algorithm (see e.g. [10]). As the starting point for the minimization, we use the observed range map $\mathbf{X}^{\mathrm{R}}$, smoothed using a median filter, and an albedo guess. The gradient of $E(\mathbf{R}, \mathbf{A})$ is computed numerically using a finite differences approximation.

2.2. Lambertian Reflectance Model

Under the Lambertian model of diffuse reflection [13], the intensity I with which a point on an object appears in the image is obtained as follows:

$$I = a \frac{\mathbf{n} \cdot \mathbf{l}}{r^2}, \tag{7}$$

where $\mathbf{n}$ is the surface normal, $\mathbf{l}$ is the unit vector from the surface point towards the light source, r is the distance of the surface point to the light source, and a is a constant that depends on the albedo of the surface, the intensity of the light source, and properties of the camera such as aperture and exposure time. For brevity, we will refer to a simply as the albedo, because any changes to a across the scene are due to albedo changes, while the properties of the light source and camera remain constant.

On a TOF camera, the light source can be assumed to be co-located with the camera, and so r is simply the range value for the surface point, and $\mathbf{l}$ is the unit vector from surface point to the camera.

2.3. Computation of Surface Normals

Some care needs to be taken when computing the surface normals on a discretely sampled range map for both

the Lambertian reflectance model and the shape prior for smooth surfaces. An obvious way is to compute the cross product of two tangent vectors $\mathbf{p}(i+1, j) - \mathbf{p}(i-1, j)$ and $\mathbf{p}(i, j+1) - \mathbf{p}(i, j-1)$ (where $\mathbf{p}(i, j)$ are the three-dimensional coordinates of the point corresponding to pixel (i, j)), but surface normals calculated in this way can lead the minimizer astray: Because the normals of pixels with even indexes depend only on the positions of pixels with odd indexes, and vice versa, neighbouring pixels are not constrained to have similar range value, and the minimizer may happily compute a surface with a "checkerboard" pattern, where neighbouring pixels are alternately displaced upwards and downwards by a certain offset, instead of forming a smooth surface.

We avoid this problem by tessellating the pixel grid into triangles. Of the many possible tessellations, we choose one where all diagonals, which divide a quad of pixels into two triangles, run parallel. Image intensity is evaluated not on the vertices of the grid but on the triangles; the intensity for a vertex is then obtained by averaging over the intensity of all triangles that are adjacent to it, weighted by the area of each triangle. The shape prior is computed over all resulting edges of the tessellation. To avoid directional artefacts, the shape prior is evaluated for two tessellations, i.e. the ones for the two possible directions of the diagonal.

2.4. Application to Time-of-Flight Cameras

When applying the method to images recorded using a TOF camera, some particular characteristics of this sensor need to be taken into account to obtain optimal results.

First, the noise σ_R in the measured range map is not the same for all pixels but depends on the amount of light collected at each pixel – the more light, the more accurate the measurement. Hence, for each pixel, we set σ_R as a function of intensity. To estimate the functional relationship between intensity and σ_R, we recorded a sequence of images of a static scene and, at each pixel, calculated the standard deviation of the measured range values. We then fit a power law function to the calculated standard deviations as a function of intensity and used this function to set σ_R in the surface reconstruction algorithm.

Another important point is that most TOF cameras do not illuminate the scene homogeneously; typically, the illumination falls off towards the edges of the field of view. To measure this effect, we recorded an image, averaged over 100 frames, of a planar object with constant albedo. By comparing the actual image $\mathbf{X}_a^I$ to the image predicted by our shading model $\mathbf{X}_p^I$ (which assumes homogeneous illumination), we were able to estimate the relative illumination strength at each pixel and use this to compensate for the ef-

fect in subsequent recordings $\mathbf{X}^I$ via

$$\mathbf{X}_{corrected}^I(i, j) = \frac{\mathbf{X}_p^I(i, j) \cdot \mathbf{X}^I(i, j)}{\mathbf{X}_a^I(i, j)}. \qquad (8)$$

Finally, if the albedo of the measured surface is approximately constant, a good initial albedo estimate can be found as follows: We find the highest-intensity pixel in the image; generally, this pixel will correspond to a part of the object that is perpendicular to the incoming light, because such regions reflect the most light. Hence, at this location, equation (7) reduces to $I = \frac{a}{r^2}$, and we obtain the albedo as $a = I r^2$. A conventional camera cannot be used to estimate albedo in this way because there, the range r is not known.

3. Results

3.1. Synthetic Data

To assess the accuracy of the method quantitatively, we first tested it on synthetic data with known ground truth. For all tests that follow, we set $w_R = 1$ and $w_A = 50$; σ_R and σ_I were set to the actual standard deviations of the noise that was added to the range map and intensity image.

Figure 1 shows the ground truth range maps for the "wave" and "corner" objects along with the noisy range map and intensity image that were used as input to the reconstruction algorithm, and the reconstruction result. The intensity image used as input to the algorithm was obtained by shading the ground truth surface with a constant albedo, then adding noise; the "global albedo" algorithm was used to reconstruct the surface. The initial albedo value for the minimization was set to twice the actual value that was used for shading. The RMS error in the reconstructed surface is reduced by a factor of around 4 for the "wave" object and almost 8 for the "corner" object.

We also measured the effect of the various components of the probabilistic model. Figure 2 shows the reconstruction error on the "wave" object as a function of the noise σ_I in the intensity image. We compare probabilistic models that use only the shading constraint $p(\mathbf{X}^I | \mathbf{R}, \mathbf{A})$, only the shape prior $p(\mathbf{R})$, or both together. (The term $p(\mathbf{X}^R | \mathbf{R})$, which incorporates the information from the measured range map, was used in all cases. Because albedo did not vary across the image, the term $p(\mathbf{A})$ was omitted.)

Not surprisingly, the error in the reconstruction based on the shading constraint increases with σ_I. The error for the shape prior does not change with σ_I because it does not use the intensity image. Finally, the error for the full probabilistic model tends to that of the shading constraint for $\sigma_I \to 0$ (because in this case, the shading term in Eq. (6) dominates the other terms) and to that of the shape prior for $\sigma_I \to \infty$ (because the shading term in Eq. (6) tends to zero). Note that the combined model yields better results than either the shading constraint or shape prior alone.
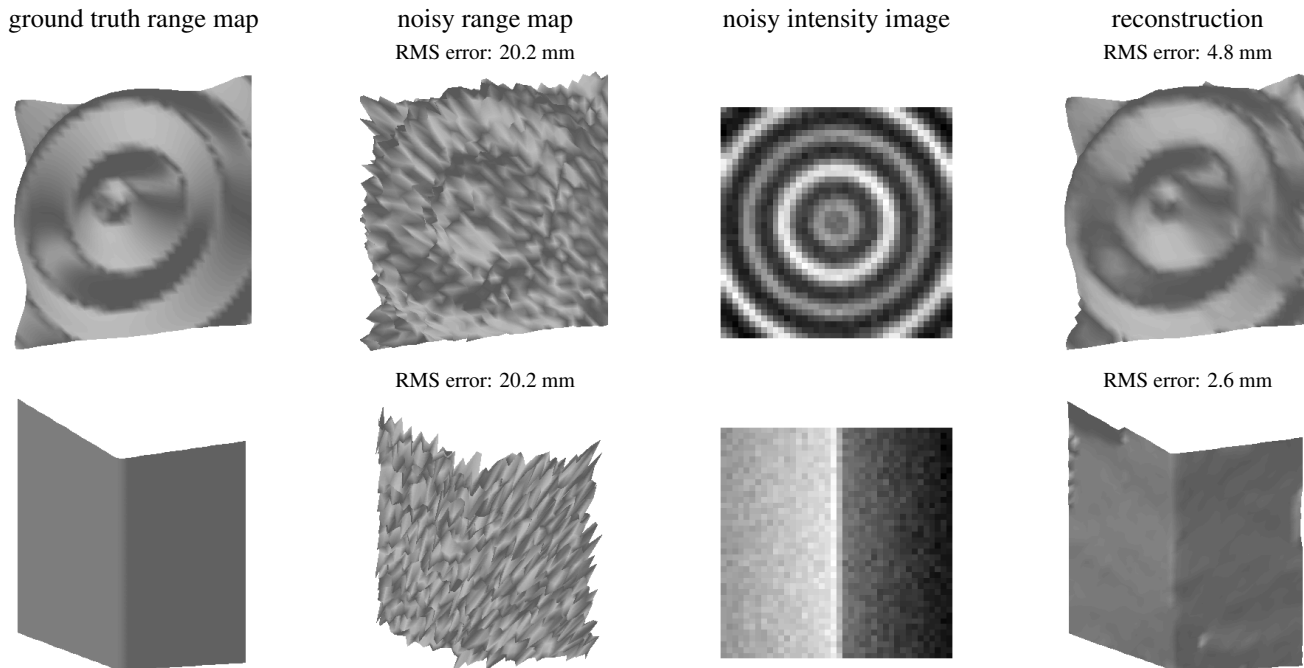
ground truth range map   noisy range map   noisy intensity image   reconstruction

RMS error: 20.2 mm   RMS error: 4.8 mm

RMS error: 20.2 mm   RMS error: 2.6 mm

Figure 1. Reconstruction results for two synthetic test objects ("wave", top, and "corner", bottom). Gaussian noise with a standard deviation of 20 mm was added to the range map; for comparison, the "wave" object has a depth of 100 mm, and the "corner" object has a depth of 120 mm. Gaussian noise with a standard deviation of 0.003 was added to the intensity image; the maximum intensity in the images was 0.19 ("corner") and 0.22 ("wave").
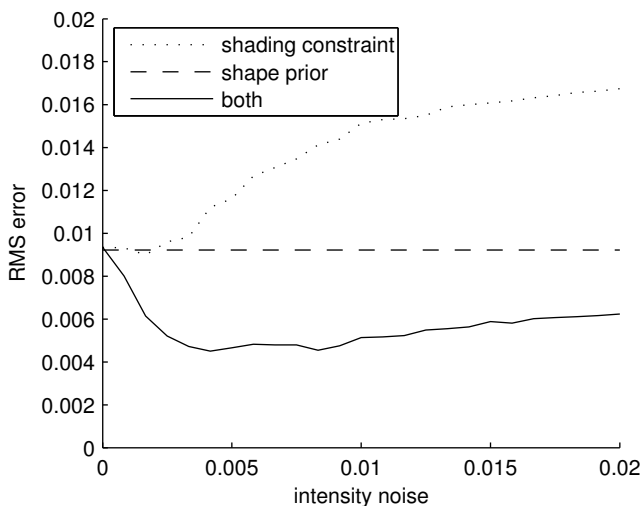

Figure 2. Reconstruction error on the "wave" object as a function of noise in the intensity image, for different probabilistic models. Range noise was fixed at a standard deviation of 20 mm.

Finally, Figure 3 shows results for the "wave" object with varying albedo. For this test, albedo was set to 0.2 on the left half of the image and 0.4 on the right half, and the noise in the range image was reduced to a standard deviation of 5 mm. Reconstructions were computed using the "global albedo" and "local albedo" algorithms. The initial albedo

value for the minimization was set to 0.3. Note that the "global albedo" algorithm does not yield a satisfactory result, while the "local albedo" version does; local albedo is estimated almost perfectly. We wish to point out, though, that albedo estimation does not always work this well; the algorithm does less well when there is less detail in the range map to "latch onto".

3.2. Real Data

We now apply the algorithm to data obtained using an SR3000 time-of-flight camera [9], which has a resolution of 176 by 144 pixels.

Figure 4 shows input data and reconstruction results for two terracotta objects. The objects were segmented manually, and the reconstruction was performed using the "global albedo" algorithm, because albedo can be assumed to be essentially constant across the object. The subjective quality of the reconstruction is greatly improved compared to the raw data; note, in particular, how the shading constraint allows us to reconstruct detail in the objects that was drowned out by noise in the measured range map.

Finally, Figure 5 shows the 3D reconstruction of a human face. We show the result of both the global and local albedo versions of the algorithm; it is evident that local albedo estimation allows a much more faithful reconstruction of areas, such as the nose, where albedo variations occur.
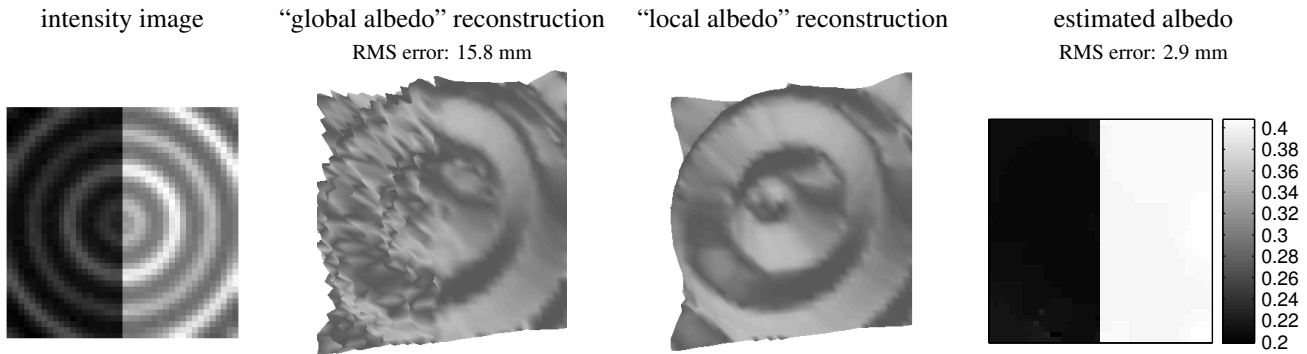
intensity image	"global albedo" reconstruction	"local albedo" reconstruction	estimated albedo
	RMS error: 15.8 mm	RMS error: 2.9 mm	

Figure 3. Reconstruction results on a "wave" object with varying albedo. The noise in the range map had a standard deviation of 5 mm, the noise in the intensity image had a standard deviation of 0.0003.
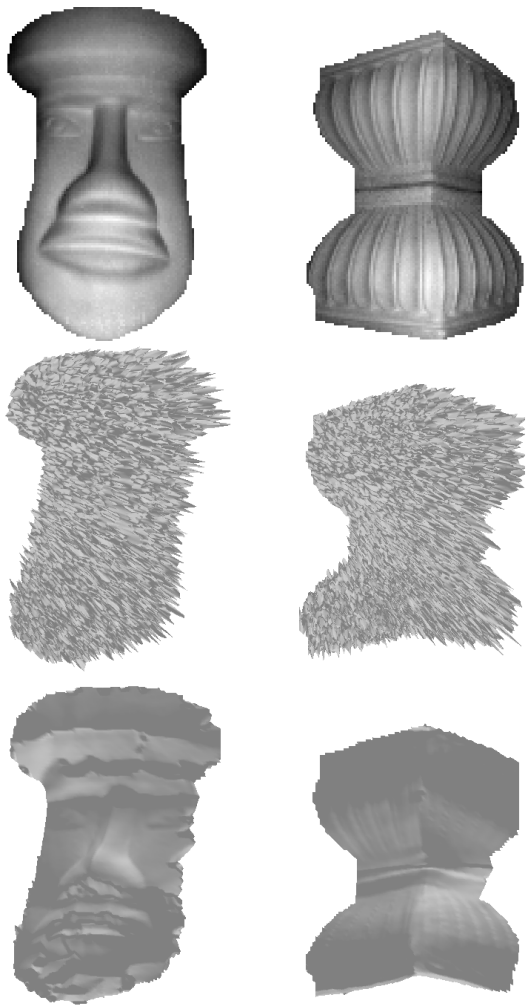

Figure 4. Surface reconstructions of two terracotta objects. The top row shows manually segmented intensity images taken with an SR3000 TOF camera. The corresponding object surfaces measured by the camera are given in the second row. The reconstructed surfaces estimated by our algorithm with "global albedo" are displayed in the bottom row.

4. Discussion

As we have shown, enforcing the shading constraint can substantially improve the quality of range maps obtained using a time-of-flight (TOF) camera, both in terms of objective measures as well as subjectively perceived quality.

The TOF camera is particularly well suited for algorithms that incorporate shape from shading (SfS) because it eliminates many sources of variability that are difficult to deal with in the general SfS setting: In the TOF camera, the position of the light source is known (it is co-located with the camera); the camera attenuates all other sources of light; and the albedo of the surface can be estimated robustly because its distance from the light source is known (see Section 2.4).

The main limitation of the current algorithm is that it does not cope well with range discontinuities, so-called *jump edges*. Because the reconstructed surface is always continuous, jump edges lead to surface normals that are almost perpendicular to the incoming light; hence, the corresponding regions are shaded with very low intensity. This disagrees with the observed image, so the algorithm will flatten the edge to compensate.

It should be possible to overcome this limitation by ignoring any mesh triangle that straddles a jump edge. Jump edges could be identified either by searching for large jumps in the measured range maps or by incorporating jump edges into the probabilistic image model, as in the work of Nadabar and Jain [8].

Of course, other range sensors, such as laser range scanners, still provide far better accuracy than TOF camera data post-processed using our algorithm. The strength of the TOF camera, however, lies in its high temporal resolution and its potential to be manufactured at low cost for mass-market applications. Enforcing the shading constraint allows TOF cameras to provide range maps of considerably enhanced quality, opening up many new application fields.

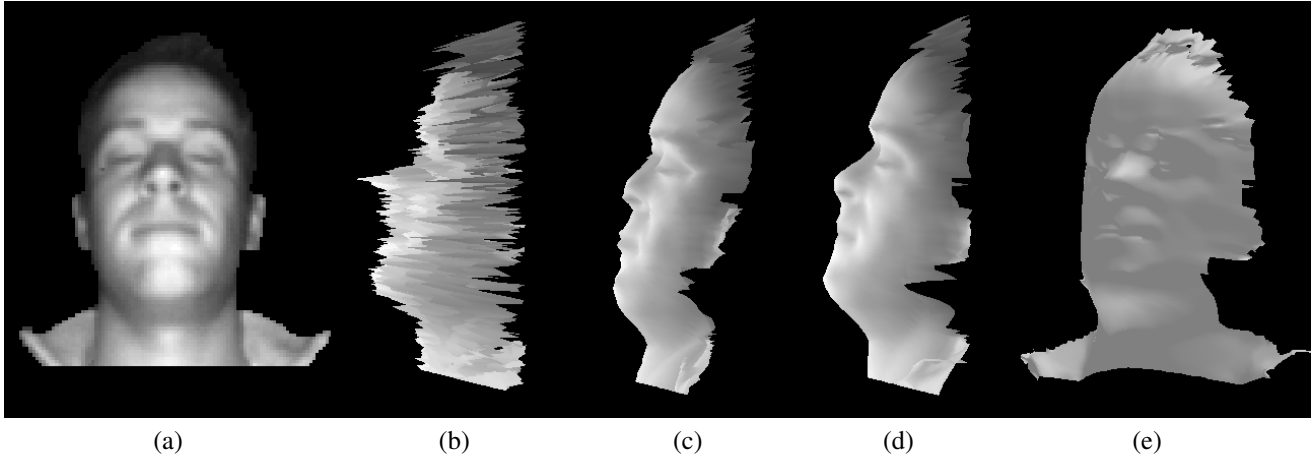<div style="text-align:center">(a)   (b)   (c)   (d)   (e)</div>

Figure 5. 3D reconstruction of a human face. The manually segmented intensity image is given in (a). A lateral view of the surface as measured by the SR3000 TOF camera is shown in (b). Figure (c) presents the reconstructed surface using the "global albedo" algorithm. Lateral and frontal views of the reconstructed surface based on the "local albedo" algorithm are shown in (d) and (e), respectively. The surface in (b), (c), and (d) was textured using the intensity image to facilitate a qualitative comparison.

References

[1] James Edwin Cryer, Ping-Sing Tsai, and Mubarak Shah. Integration of shape from shading and stereo. *Pattern Recognition*, 28(7):1033–1043, 1995.

[2] James R. Diebel and Sebastian Thrun. An application of Markov random fields to range sensing. In *Advances in Neural Information Processing Systems 18*, pages 291–298, 2006.

[3] James R. Diebel, Sebastian Thrun, and Michael Brünig. A Bayesian method for probable surface reconstruction and decimation. *ACM Transactions on Graphics*, 25(1):39–59, 2006.

[4] Jean-Denis Durou, Maurizio Falcone, and Manuela Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43, 2008.

[5] Pascal V. Fua and Yvan G. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *International Journal of Computer Vision*, 16(1):35–56, 1995.

[6] Yvan G. Leclerc and Aaron F. Bobick. The direct computation of height from shading. In *Computer Vision and Pattern Recognition (CVPR '91)*, pages 552–558, 1991.

[7] Mostafa G.-H. Mostafa, Sameh M. Yamany, and Aly A. Farag. Integrating shape from shading and range data using neural networks. In *Computer Vision and Pattern Recognition (CVPR '99)*, volume 2, page 2015, 1999.

[8] Sateesha G. Nadabar and Anil K. Jain. Fusion of range and intensity images on a Connection Machine (CM-2). *Pattern Recognition*, 28(1):11–26, 1995.

[9] Thierry Oggier, Bernhard Büttgen, Felix Lustenberger, Guido Becker, Björn Rüegg, and Agathe Hodac. SwissRanger^{TM} SR3000 and first experiences based on miniaturized 3D-TOF cameras. In *Proceedings of the 1st Range Imaging Research Day*, pages 97–108, Zürich, Switzerland, 2005.

[10] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, second edition, 1992.

[11] Dimitrios Samaras, Dimitris Metaxas, Pascal V. Fua, and Yvan G. Leclerc. Variable albedo surface reconstruction from stereo and shape from shading. In *Proceedings of Computer Vision and Pattern Recognition*, volume 1, pages 480–487, 2000.

[12] Clay Matthew Thompson. Robust photo-topography by fusing shape-from-shading and stereo. AI Technical Report 1411, Massachusetts Institute of Technology, 1993.

[13] Emanuele Trucco and Alessandro Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1998.