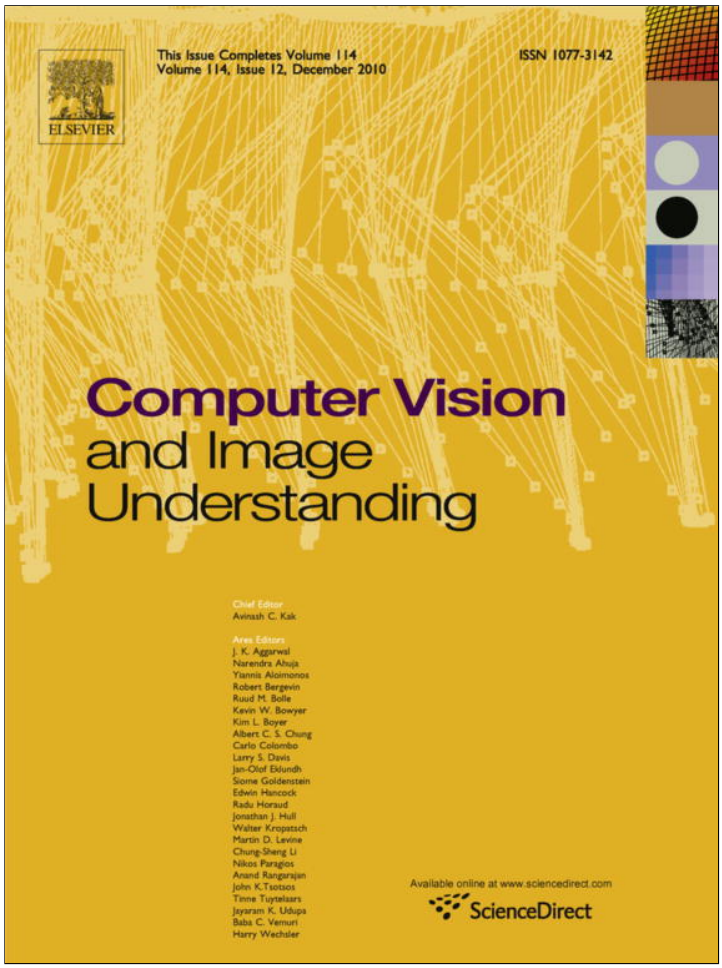

Computer Vision and Image Understanding 114 (2010) 1329–1335

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Shading constraint improves accuracy of time-of-flight measurements

Martin Böhme *, Martin Haker, Thomas Martinetz, Erhardt Barth

Institute for Neuro- and Bioinformatics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

ARTICLE INFO

Article history:
Received 5 January 2009
Accepted 3 August 2010
Available online 15 September 2010

Keywords:
Time-of-flight camera
Range map
Range sensor
Shading constraint
Shape from shading
Probabilistic image model

ABSTRACT

We describe a technique for improving the accuracy of range maps measured by a time-of-flight (TOF) camera, a novel type of sensor that provides a range map registered perfectly with an intensity image.

Our 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. In practice, a general reflectance model (such as Lambertian reflectance) provides a sufficient approximation for a wide range of surfaces.

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. We also show how the reflectivity (or *albedo*) of the surface can be estimated, both globally for an entire object and locally for objects where albedo varies across the surface.

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 [18,4] for surveys). 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 places a strong additional constraint on the shape of the object, allowing ambiguities that may exist in the pure SfS setting [5] to be resolved and enabling the albedo of the surface to be estimated, both globally for an entire object as well as locally for objects where albedo varies across the surface.

Besides the shading constraint, there are also other ways of 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 [12] use a Markov random field (MRF) to identify different types of edges. Diebel and Thrun [2] use edge strengths estimated on a high-resolution color image to increase the resolution of a low-resolution depth map.

The idea of integrating the shading constraint with other range information has been investigated by a number of researchers. Most of this work focuses on the integration of SfS with stereo. These two techniques complement each other well because SfS works well on uniformly colored areas, whereas stereo requires surface texture to find stereo correspondences. Because of this, the fusion of SfS with stereo has a slightly different focus than the fusion of SfS with a TOF range map. In the stereo case, we only have range information in textured areas and need to rely on shading cues in untextured areas. In the TOF case, we have a dense range map and wish to fuse information from TOF and shading at the same pixel.

Many approaches to the fusion of SfS and stereo (see for example the work of Thompson [16], Fua and Leclerc [6], and Hartt and Carlotto [9]) use an objective function that depends directly on the two or more images obtained from a multi-camera setup; for this reason, they do not generalize to settings where a range map has been obtained in some other way than through stereo. Samaras et al. [15] combine stereo with SfS by using stereo in textured areas

* Corresponding author.
E-mail address: boehme@inb.uni-luebeck.de (M. Böhme).

1077-3142/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2010.08.001

and SfS in untextured areas, but they do not perform a fusion of stereo and SfS at the same location. Haines and Wilson [7,8] fuse stereo and SfS in a probabilistic approach based on a disparity map and the shading observed in one of the stereo images. Because there is a one-to-one correspondence between disparity and range, the approach could also be used with range maps obtained by arbitrary means. However, since color is used to segment areas of different albedo, the approach is not suitable for use with TOF cameras, which typically only deliver a grayscale image.

There are several other approaches that 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 [10] 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. [11] 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: First of all, we assume that $p(\mathbf{X}^{\mathrm{R}}, \mathbf{X}^{\mathrm{I}} | \mathbf{R}, \mathbf{A})$ can be written as follows:

$$p(\mathbf{X}^{\mathrm{R}}, \mathbf{X}^{\mathrm{I}} | \mathbf{R}, \mathbf{A}) = p(\mathbf{X}^{\mathrm{R}} | \mathbf{R}, \mathbf{A}) p(\mathbf{X}^{\mathrm{I}} | \mathbf{R}, \mathbf{A}). \tag{2}$$

In other words, the observations $\mathbf{X}^{\mathrm{R}}$ and $\mathbf{X}^{\mathrm{I}}$ are conditionally independent given $\mathbf{R}$ and $\mathbf{A}$.

We now assume that the observed range map $\mathbf{X}^{\mathrm{R}}$ is simply the true range map $\mathbf{R}$ with additive Gaussian noise, i.e.

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

Note that the standard deviation σ_{R} is not constant but can vary per pixel as a function of range and albedo. As we will see in Section 2.4, the noise in the range measurement of a TOF camera depends on the amount of light that returns to the camera.

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.

$$p(\mathbf{X}^{\mathrm{I}} | \mathbf{R}, \mathbf{A}) = \mathcal{N}(\mathbf{X}^{\mathrm{I}} - \mathbf{I}(\mathbf{R}, \mathbf{A}) | \mu = 0, \sigma_{\mathrm{I}}). \tag{4}$$

For the range map prior $p(\mathbf{R})$, we use the shape prior introduced by Diebel et al. [3], which favors surfaces with smoothly changing surface normals. We tessellate the range map into triangles (see Section 2.3) and compute the surface normal $\mathbf{n}_j$ for each triangle. The shape prior is then given by the energy function

$$E^{\mathrm{R}}(\mathbf{R}) = w_{\mathrm{R}} \sum_{\substack{\text{triangles } j,k \\ \text{adjacent}}} \|\mathbf{n}_j - \mathbf{n}_k\|_2, \tag{5}$$

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) "Fixed albedo": A single albedo value, specified beforehand, is used for all pixels. (ii) "Global albedo": 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) "Local albedo": Each pixel location may have a different albedo, and the prior $p(\mathbf{A})$ favors smooth albedo changes. In this latter case, we use an energy function

$$E^{\mathrm{A}}(\mathbf{A}) = w_{\mathrm{A}} \sum_{\substack{\text{pixels } j,k \\ \text{adjacent}}} |a_j - a_k|, \tag{6}$$

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 posterior and eliminate constant additive terms to obtain an energy function

$$E(\mathbf{R}, \mathbf{A}) = \sum_j \frac{\left(X_j^{\mathrm{R}} - R_j\right)^2}{2\sigma_{\mathrm{R}}^2} + \sum_j \frac{\left(X_j^{\mathrm{I}} - I_j(\mathbf{R}, \mathbf{A})\right)^2}{2\sigma_{\mathrm{I}}^2} + \quad E^{\mathrm{R}}(\mathbf{R}) + E^{\mathrm{A}}(\mathbf{A}), \tag{7}$$

where the index j runs over all pixels. (For the "fixed albedo" and "global albedo" models, the term $E^{\mathrm{A}}(\mathbf{A})$ is omitted.) Note that all the terms in the energy function are unitless due to multiplication or division by the constants σ_{R}, σ_{I}, w_{R}, and w_{A}.

We find the maximum a posteriori 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. [14]). 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 (see Section 2.4). The gradient of $E(\mathbf{R}, \mathbf{A})$ is computed numerically using a finite differences approximation. The parameters σ_{R}, σ_{I}, w_{R}, and w_{A} should be set to reflect the noise characteristics of the sensor and the statistical properties of the scene.

2.2. Lambertian reflectance model

Under the Lambertian model of diffuse reflection [17], the intensity I with which a point on an object appears in the image is obtained as follows:

$$I = a \frac{\mathbf{n} \cdot \mathbf{l}}{r^2}, \tag{8}$$

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 the surface point to the camera.

2.3. Computation of surface normals

Both the Lambertian reflectance model and the shape prior for smooth surfaces require the normals of the surface to be known; some care needs to be taken when computing these normals on a discretely sampled range map. 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 indices depend only on the positions of pixels with odd indices, and vice versa, neighboring pixels are not constrained to have similar range values, and the minimizer may happily compute a surface with a "checkerboard" pattern, where neighboring pixels are alternately displaced upwards and downwards by a certain offset, instead of forming a smooth surface.

For this reason, care needs to be taken when evaluating the reflectance model and shape prior to obtain a formulation that is physically realistic and does not lead the minimizer astray. For any given pixel, the range and intensity measured by the camera are averaged over the area covered by the pixel. Nevertheless, we will assume that these values correspond to the range and intensity of an individual point on the object's surface. To obtain a continuous surface between this grid of points, we tessellate the grid into triangles. Of the many possible tessellations, we choose one where the diagonals dividing a quad of pixels into two triangles run parallel.

To compute the intensity of a pixel, given a certain range map, we compute the average intensity over all triangles adjacent to it. (All triangles have the same projected area in the image, hence they are all weighted equally.) Because the triangles in the mesh are small compared to their distance from the camera, we can assume that intensity is constant across a triangle. The intensity I_j for pixel j is thus obtained as follows:

$$I_j = \frac{a_j \sum_{k \in N_j} \mathbf{n}_k \cdot \mathbf{l}_j}{R_j^2 |N_j|}, \tag{9}$$

where a_j is the albedo of the pixel, R_j is the range value of the pixel, $\mathbf{l}_j$ is the unit vector from the surface point to the light source, $\mathbf{n}_k$ is the surface normal of triangle k, and N_j is the set of triangles that are adjacent to pixel j; see Fig. 1 for an illustration of the triangles that are adjacent to a pixel.

When computing the shape prior, we must take care to count each edge exactly once. We do this by iterating over all pixels and, for each pixel, evaluating the shape prior only for those edges in the tessellation that lie below and to the right of the pixel in the grid. This is illustrated in Fig. 2a, where the bold edges are the ones for which the shape prior is evaluated when considering the central pixel.

The direction in which the diagonals run in the tessellation introduces an asymmetry, and we have found that this asymmetry can cause the shape prior to generate directional artefacts. For this reason, we evaluate the shape prior for both of the two possible directions of the diagonal and sum over the results. Fig. 2b shows which edges are evaluated for a given pixel on the tessellation with the alternative direction of the diagonal.

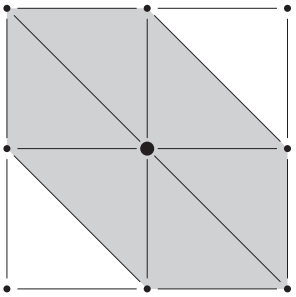

Fig. 1. The intensity of a pixel is computed by averaging over the intensity of all triangles that are adjacent to it. Triangles that are adjacent to the central pixel are shaded in gray.

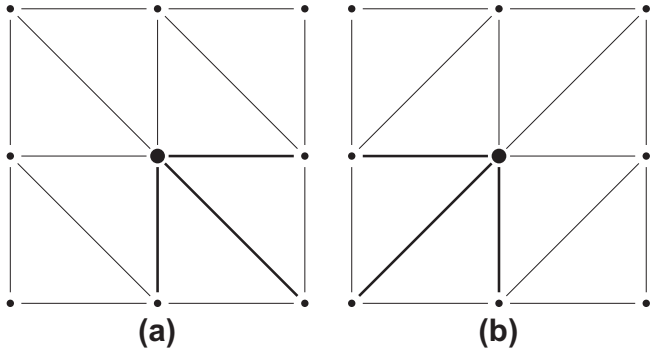

Fig. 2. To avoid directional artefacts, the shape prior is evaluated over two different tessellations, with the diagonals running in opposite directions. (a) Tessellation with diagonals running top left to bottom right. For each pixel, the shape prior is evaluated for the edges (shown in bold) below and to the right of the pixel. (b) Tessellation with diagonals running bottom left to top right. The shape prior is evaluated for the edges below and to the left of the pixel.

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

$$\mathbf{X}^I_{\text{corrected}}(i,j) = \frac{\mathbf{X}^I_p(i,j) \cdot \mathbf{X}^I(i,j)}{\mathbf{X}^I_a(i,j)}. \tag{10}$$

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, Eq. (8) reduces to $I = \frac{a}{r^2}$, and we obtain the albedo as $a = Ir^2$. 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: A rotationally symmetric sinusoid (the "wave" object) and an object composed of two planar surfaces that meet at a sharp edge (the "corner" object); see Fig. 3. To simulate the measurement process of the TOF camera, we shaded the ground truth surface with a constant albedo, then added Gaussian noise; the observed range map was obtained by adding Gaussian noise to the ground truth surface. 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.

Fig. 3 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. For comparison, the figure also shows the result of filtering the range map with a 5×5 median filter. The noise in the range map had a standard deviation of 20 mm; for comparison, the "wave" object has a depth of 100 mm, and the "corner" object has a depth of 120 mm. The intensity image noise had a standard deviation of 0.003; the maximum intensity in the images was 0.19 ("corner") and 0.22 ("wave"). 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 to produce the intensity image. The RMS error in the reconstructed surface is reduced by a factor of over 4 for the "wave" object and around 8 for the "corner" object.

Next, we examine the results of the algorithm on an object with varying albedo. First, we use the "wave" object with albedo set to 0.2 on the left half of the image and 0.4 on the right half (Fig. 4, top); 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. A second test (Fig. 4, middle) used the same object but with albedo varying continuously from 0.2 at the left to 0.4 at the right. Albedo is overestimated slightly on the left side of the image, and the result is not quite as good as in the first case but still satisfactory. Finally, we show a case where the albedo estimation does not work properly (Fig. 4, bottom): the "corner" object with albedo varying continuously between 0.2 at the top and 0.4 at the bottom. Here, the result of the "local albedo" algorithm is not satisfactory and, in fact, its RMS error is higher than that of the "global albedo" algorithm. We suspect the reason for the poor performance may be that the range map does not contain enough detail for the algorithm to "latch onto".

Finally, we measured the effect of the various components of the probabilistic model. Fig. 5 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 $p(\mathbf{X}^R|\mathbf{R},\mathbf{A})$, 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.)

Since σ_I appears in the denominator of the shading term in the energy function (Eq. (7)), we need to treat $\sigma_I = 0$ (no intensity noise) as a special case. Note that for $\sigma_I \to 0$, the shading term dominates all the other terms; hence, if $\sigma_I = 0$, we omit all other terms from the energy function. We can then avoid scaling by $\frac{1}{\sigma_I^2}$.

The error for the reconstruction obtained using only the shape prior (along with the measured range map $p(\mathbf{X}^R|\mathbf{R},\mathbf{A})$) is, of course, constant for all σ_I because it does not use the intensity image. The shape prior reduces the RMS error in the range map by around a factor of 2.

For the reconstructions obtained using only the shading constraint (along with the measured range map), the error in the

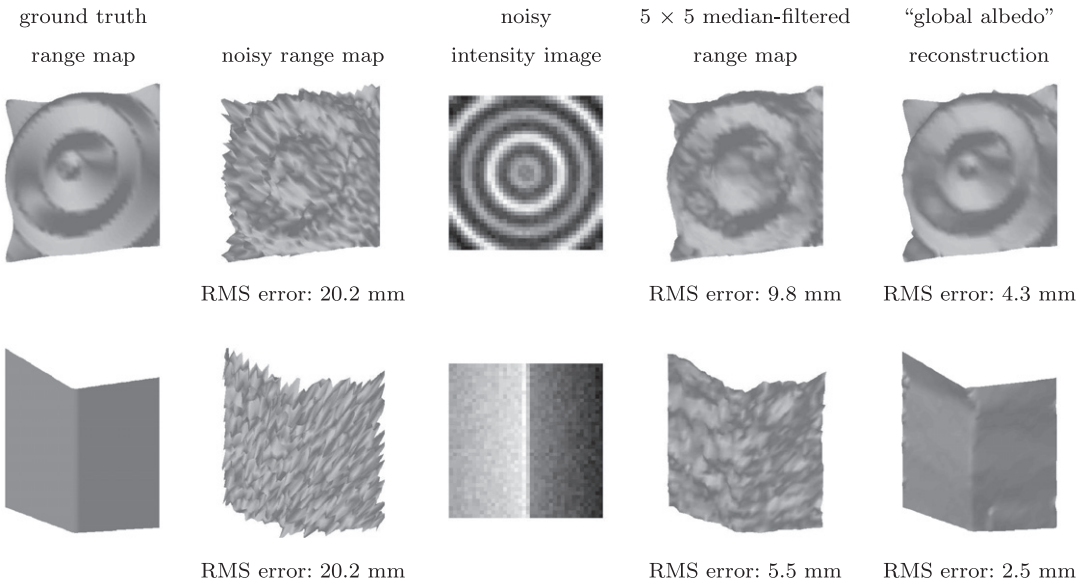

Fig. 3. Reconstruction results for two synthetic test objects ("wave", top, and "corner", bottom). Gaussian noise with a standard deviation of 20 mm was added to the range map; for comparison, the "wave" object has a depth of 100 mm, and the "corner" object has a depth of 120 mm. Gaussian noise with 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").

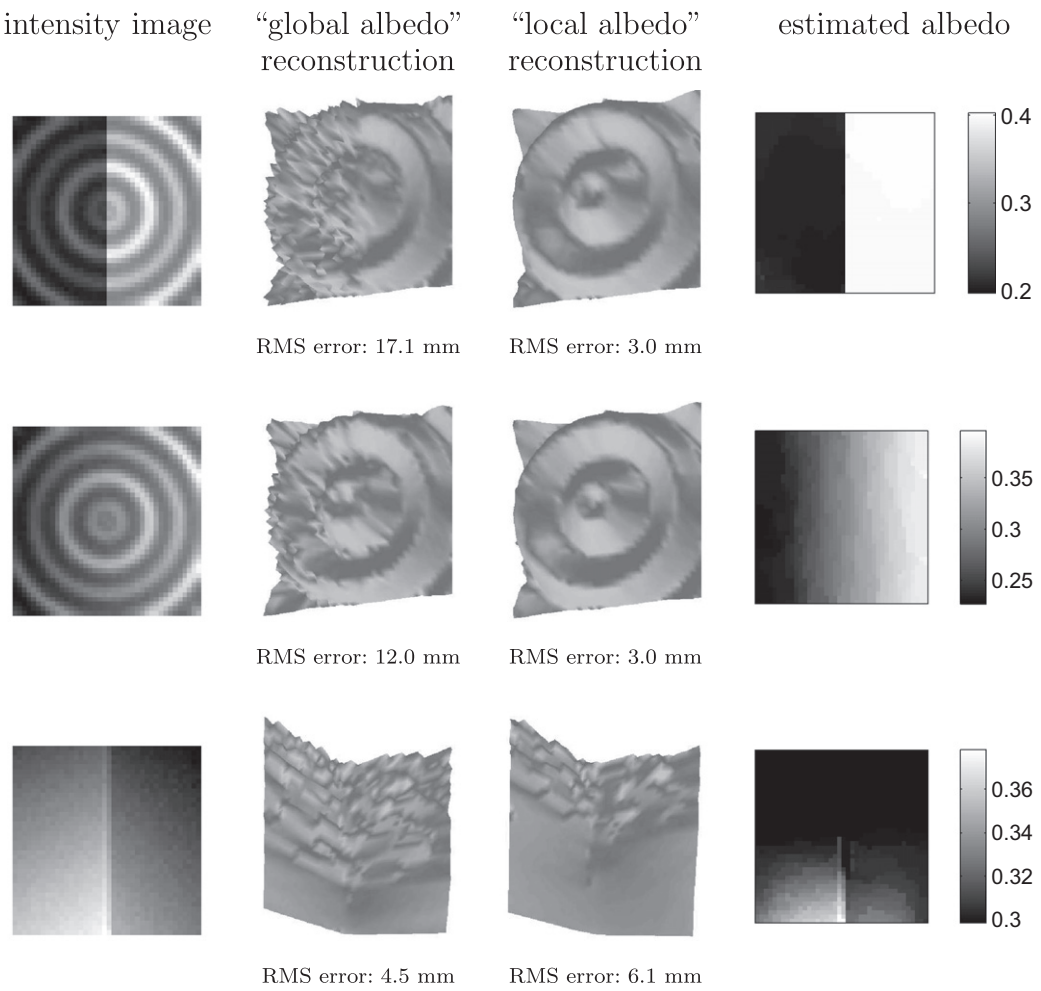

Fig. 4. Reconstruction results on synthetic objects with varying albedo. Top: "wave" object with an albedo of 0.2 on the left half of the object and 0.4 on the right half. Middle: "wave" object with albedo varying continuously from 0.2 at the left to 0.4 at the right. Bottom: "Corner" object with albedo varying continuously from 0.2 at the top to 0.4 at the bottom. In all cases, the noise in the range map had a standard deviation of 5 mm, and the noise in the intensity image had a standard deviation of 0.003.

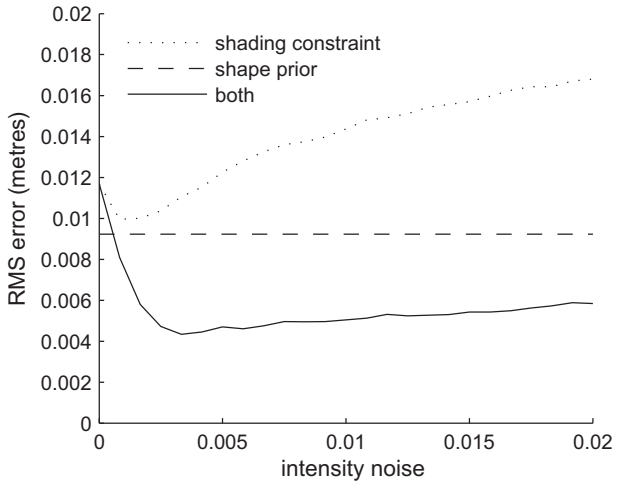

Fig. 5. Reconstruction error on the "wave" object as a function of noise in the intensity image, for different probabilistic models. (For comparison, the maximum intensity in the image was 0.22.) Range noise was fixed at a standard deviation of 20 mm.

reconstruction generally increases with the noise σ_I in the intensity image. However, it is notable that, even for $\sigma_I = 0$, the range map is not reconstructed perfectly. Though, in this case, the ground truth range map is obviously the global minimum of the energy function, the algorithm appears to get stuck in a local minimum. Recall that in the case $\sigma_I = 0$, the shading term dominates all other terms in the energy function. As σ_I increases, the energy function begins taking the measured range map into account, and this in fact leads to an initial slight reduction in the reconstruction error; we speculate that the additional constraint imposed by the measured range map makes it easier to minimize the energy function.

This effect is even more prononunced for the full model, which combines the shading constraint, the shape prior, and the measured range map. For $\sigma_I = 0$, the shading term again dominates all other terms in the energy function, and so we obtain the same result as for the shading constraint alone. As σ_I begins to increase, the reconstruction error decreases markedly as the shape prior and the measured range map come into play. After a certain point, the reconstruction error begins increasing again; for $\sigma_I \to \infty$, the reconstruction error will tend to that of the range prior because the shading term in Eq. (7) tends to zero. Note that, except for very small σ_I, the combined model yields better results than either the shading constraint or the shape prior alone.

3.2. Real-world data

We now apply the algorithm to data obtained using an SR3000 time-of-flight camera [13], which has a resolution of 176 by 144

pixels. The parameters σ_R and σ_I (standard deviations of range and intensity) were set to values estimated on a sequence of images of a static scene; σ_R was set as a function of intensity for each pixel (see Section 2.4), while σ_I was constant across the whole scene, in accordance with the statistical properties of the sensor. The parameters for the shape and albedo prior were again set to $w_R = 1$ and $w_A = 50$.

We first demonstrate the algorithm on two terracotta objects, which fulfill the assumption of Lambertian reflectance quite well and can be assumed to have approximately constant albedo. Fig. 6 shows the input data and reconstruction results for the two terracotta objects. To compare the results with the effect that a conventional filtering has on the range map, a 5×5 median-filtered version is also shown. The objects were segmented manually, and the reconstruction was performed using the "global albedo" algorithm. The subjective quality of the reconstruction is greatly improved compared to both the raw data and the median-filtered version; 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.

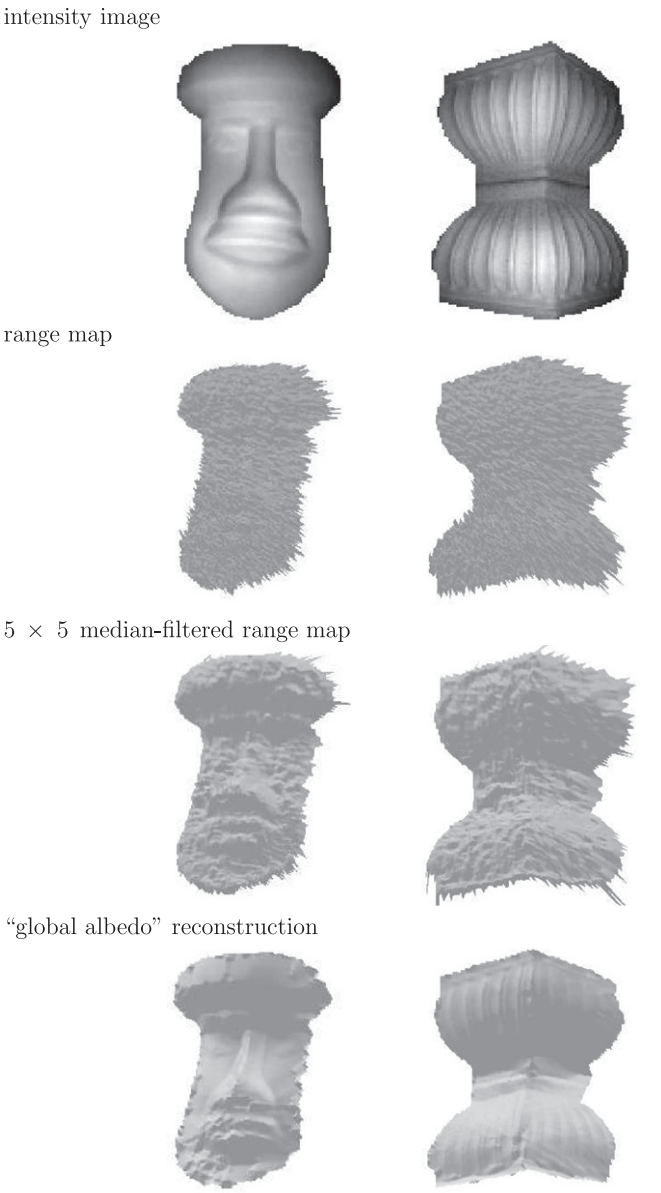

Fig. 6. Surface reconstructions of two terracotta objects, manually segmented in images taken with an SR3000 TOF camera. The renderings of the range maps are rotated 30° around the vertical axis.

Fig. 7 shows the results of the algorithm on a human face. This is a more challenging object for the algorithm because the reflectance properties of skin are considerably more complex than the Lambertian reflectance assumed by the model; also, albedo variations occur in places such as the eyebrows and the lips.

We show the result of both the global and local albedo versions of the algorithm; again, a 5×5 median-filtered version is also shown. It is evident that local albedo estimation allows a much more faithful reconstruction than global albedo estimation in areas, such as the lips, where albedo variations occur. Also, despite the fact that skin is not a Lambertian reflector, the shape of the face is reconstructed quite accurately, demonstrating that the algorithm is not very sensitive to violations of the Lambertian reflectance assumption.

Finally, Fig. 8 shows the results of the algorithm on the upper body of a person. Note how the shading constraint allows the cloth folds to be reconstructed faithfully. This example also illustrates the limitations of the algorithm: The head is reconstructed less well than in the previous example; we believe this is because there is too much albedo variation in an area of only a few pixels. The lowest part of the body is not reconstructed well either, and this is probably due to the low reflectivity of the material in this region, which leads to a large amount of noise in both the range map and the intensity image.

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 [12].

It should also be noted that the algorithm is computationally fairly expensive; the current implementation takes several minutes to process an image on a contemporary PC. Since our main focus was correctness, not performance, we expect that optimization should yield a substantial speedup. However, optimizing the algorithm to the point that it could run at camera frame rates would present a major challenge and would probably require techniques such as computation on the graphics processing unit (GPU).

Even in its present form, though, the algorithm is suitable for the post-processing either of individual images or of recorded image sequences. Of course, other range sensors, such as laser range scanners, still provide far better accuracy than TOF camera data

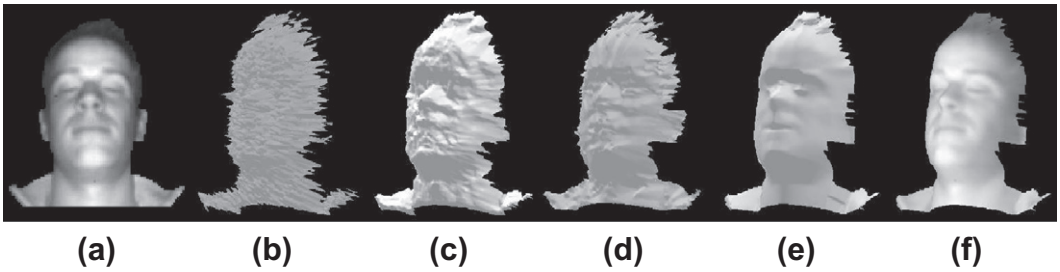

Fig. 7. 3D reconstruction of a human face. (a) Manually segmented intensity image, (b) measured range map, (c) 5×5 median-filtered range map, (d) "global albedo" reconstruction, (e) "local albedo" reconstruction, and (f) "local albedo" reconstruction textured with intensity image. The renderings of the range maps are rotated 30 degrees around the vertical axis.

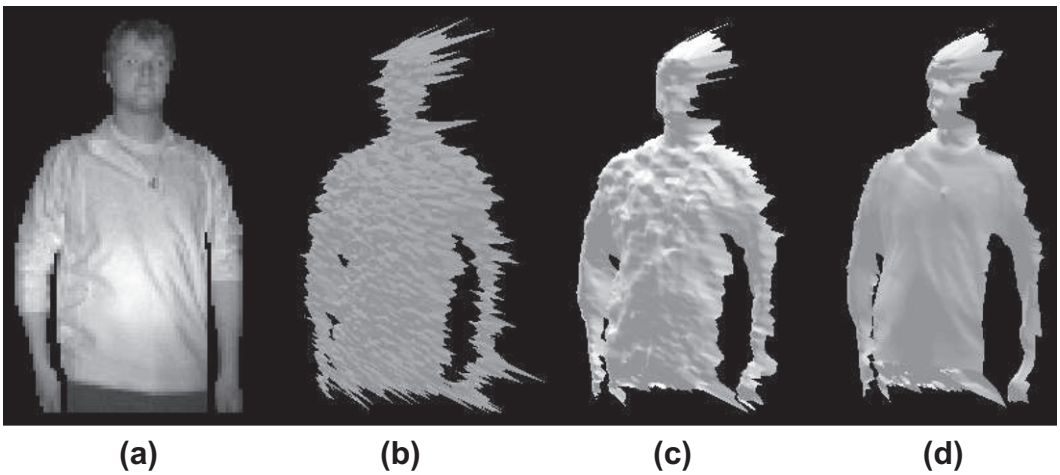

Fig. 8. 3D reconstruction of a person's upper body. (a) Manually segmented intensity image, (b) measured range map, (c) 5×5 median-filtered range map, and (d) "local albedo" reconstruction. The renderings of the range maps are rotated 30° around the vertical axis.

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.

Acknowledgments

We thank the anonymous reviewers for their helpful remarks. This work was developed within the ARTTS project (www.artts.eu), which is funded by the European Commission (contract no. IST-34107) within the Information Society Technologies (IST) priority of the 6th Framework Programme. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References

[1] J.E. Cryer, P.-S. Tsai, M. Shah, Integration of shape from shading and stereo, Pattern Recognition 28 (7) (1995) 1033–1043.
[2] J.R. Diebel, S. Thrun, An application of Markov random fields to range sensing, in: Advances in Neural Information Processing Systems, vol. 18, 2006.
[3] J.R. Diebel, S. Thrun, M. Brünig, A Bayesian method for probable surface reconstruction and decimation, ACM Transactions on Graphics 25 (1) (2006) 39–59.
[4] J.-D. Durou, M. Falcone, M. Sagona, Numerical methods for shape-from-shading: a new survey with benchmarks, Computer Vision and Image Understanding 109 (1) (2008) 22–43.
[5] J.-D. Durou, D. Piau, Ambiguous shape from shading with critical points, Journal of Mathematical Imaging and Vision 12 (2) (2000) 99–108.
[6] P.V. Fua, Y.G. Leclerc, Object-centered surface reconstruction: combining multi-image stereo and shading, International Journal of Computer Vision 16 (1) (1995) 35–56.
[7] T.S.F. Haines, R.C. Wilson, Integrating stereo with shape-from-shading derived orientation information, in: British Machine Vision Conference, vol. 2, 2007.
[8] T.S.F. Haines, R.C. Wilson, Combining shape-from-shading and stereo using Gaussian-Markov random fields, in: International Conference on Pattern Recognition, 2008.
[9] K. Hartt, M. Carlotto, A method for shape-from-shading using multiple images acquired under different viewing and lighting conditions, in: Proceedings of Computer Vision and Pattern Recognition, 1989.
[10] Y.G. Leclerc, A.F. Bobick, The direct computation of height from shading, in: Computer Vision and Pattern Recognition (CVPR '91), 1991.
[11] M.G.-H. Mostafa, S.M. Yamany, A.A. Farag, Integrating shape from shading and range data using neural networks, in: Computer Vision and Pattern Recognition (CVPR '99), vol. 2, 1999.
[12] S.G. Nadabar, A.K. Jain, Fusion of range and intensity images on a Connection Machine (CM-2), Pattern Recognition 28 (1) (1995) 11–26.
[13] T. Oggier, B. Büttgen, F. Lustenberger, G. Becker, B. Rüegg, A. Hodac, SwissRanger™ SR3000 and first experiences based on miniaturized 3D-TOF cameras, in: Proceedings of the 1st Range Imaging Research Day, Zürich, Switzerland, 2005.
[14] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C, second ed., Cambridge University Press, Cambridge, UK, 1992.
[15] D. Samaras, D. Metaxas, P.V. Fua, Y.G. Leclerc, Variable albedo surface reconstruction from stereo and shape from shading, in: Proceedings of Computer Vision and Pattern Recognition, vol. 1, 2000.
[16] C.M. Thompson, Robust photo-topography by fusing shape-from-shading and stereo, AI Technical Report 1411, Massachusetts Institute of Technology, 1993.
[17] E. Trucco, A. Verri, Introductory Techniques for 3-D Computer Vision, Prentice Hall, 1998.
[18] R. Zhang, P.-S. Tsai, J.E. Cryer, M. Shah, Shape from shading: a survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (8) (1999) 690–706.