A Benchmarking Dataset for Performance Evaluation of Automatic Surface Reconstruction Algorithms

Anke Bellmann, Olaf Hellwich, Volker Rodehorst and Ulaş Yılmaz

Computer Vision & Remote Sensing, Berlin University of Technology
Franklinstr. 28/29, FR 3-1, 10587 Berlin, Germany
{bellmann, hellwich, vr, ulas}@cs.tu-berlin.de

http://www.cv.tu-berlin.de/

Abstract

Numerous techniques were invented in computer vision and photogrammetry to obtain spatial information from digital images. We intend to describe and improve the performance of these vision techniques by providing test objectives, data, metrics and test protocols. In this paper we propose a comprehensive benchmarking dataset for evaluating a variety of automatic surface reconstruction algorithms (shape-from-X) and a methodology for comparing their results.

1. Introduction

The goal of this study is to investigate various methods for automatic reconstruction of 3D surfaces from a collection of images. In computer vision such a task is called shape-from-*X*, where *X* denotes the cue used to infer shape, for example

- shape-from-stereo
- shape-from-motion
- shape-from-silhouette
- shape-from-texture
- shape-from-shading
- shape-from-(de-)focus

Although many robust algorithms are already available in the literature and in practice, there exist hardly any methods to test and compare the performance of these algorithms quantitatively [3]. We are aware that this task seems to be very general and the lowest common denominator is small. Particularly, the algorithms have different properties, overall characteristics and rely on different assumptions. Nevertheless, in our opinion a methodology for evaluating surface reconstruction algorithms is important and still at its beginning [23].

In this study we established a true benchmark dataset for the performance evaluation of shape-from-*X* algo-rithms. The experimental setup is challenging due to controversial requirements of these surface reconstruction algorithms. For instance, the reflectance based methods [9, 27] deal with curved Lambertian surfaces, whereas image matching approaches [1, 16, 18] in general prefer textured and/or piecewise planar objects. Therefore, we used a combination of diffuse objects with different surface geometries and synthetic projected textures.

We also propose a method for automatically evaluating the results obtained through these algorithms. A comparative study is difficult due to various reasons. For instance, binocular stereo produces dense depth or disparity maps of the object, whereas multi-view stereo reconstructs the object surface as a 3D polygon mesh or unstructured point cloud and finally reflectance-based methods give surface orientations instead of depth information. Hence, we decided to compare 2.5D reconstruction results from one reference view to obtain a measure of success and a ranking.

However, the ranking should not be considered as a direct measure of success of one method over another. Instead, it allows an objective comparison of different approaches and provides an informative basis for the combination of reconstruction methods. The limitations of each method are already surveyed and can be found in the literature. However, our aim in establishing this dataset is to provide researchers a tool, by which they can see how successful or unsuccessful their method with respect to other methods is.

The organization of the paper is as follows. In the following section, the established benchmarking dataset is explained in detail. An analysis of the surface reconstruction algorithms, which are under consideration, is given in Section 3. The evaluation methodology and ranking strategies are discussed in Section 4. The paper concludes with Section 5 in which we discuss the presented study and state possible improvements.

2. Benchmarking Data

In this section, we describe how we created the benchmarking dataset, which consists of

- 360 color images of a real scene,
- 20 synthetic rendered scene images,
- 4 real and 38 synthetic texture patterns,
- orientation and calibration data for
  - 52 camera positions,
  - a texture-mapping LCD-projector,
  - 3 light source directions,
- ground truth in terms of
  - a 3D point cloud,
  - a depth map,
  - a surface orientation map.

2.1. Object Selection

Finding a single representative object that satisfies the requirements of all reconstruction algorithms is the first challenge in creating the benchmarking data. The surface of a representative object should consist of smooth curved parts as well as piecewise planar patches. On one hand it should have a Lambertian surface and on the other hand enough textural information on it. Therefore, we have assembled various objects with different surface geometries, in order to obtain the benchmarking scene with moderate occlusions (see Figure 1).

The object surface is made of white plaster, which approximate a good Lambertian surface with perfectly matte reflectance properties. Texture, which is necessary to solve the correspondence problem, is projected using an LCD-projector. This allows analyzing the effect of different texture patterns on the final result. Unfortunately, artificial shadows are introduced that penalize texture-based approaches. As a result, the final scene is structured enough to rank success of different reconstruction results, as well as general enough to be used by a large variety of algorithms.
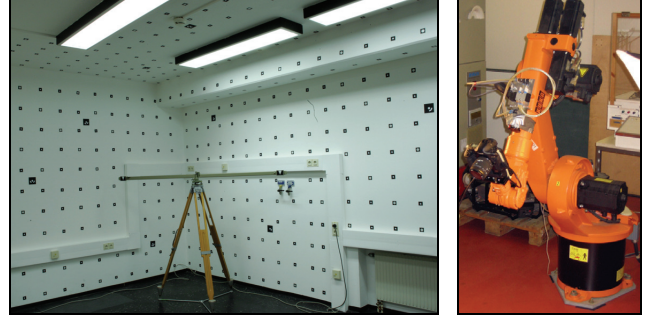

Figure 1: The benchmarking scene object

Figure 2:   Control point field for camera calibration and the
KUKA robot arm with the mounted digital camera

2.2. Calibration

2.2.1   Camera

Images are captured using the digital single-lens reflect camera *Canon EOS-1D Mark II*, which has a 50 mm lens and a 28.7×19.1 mm CMOS sensor. Maximum image size is 3504×2336 pixels. For our experiments this resolution is reduced to 1728×1152 and the captured images are cropped to a region of interest with 1360×904 pixels.

Interior orientation of the camera is computed with the bundle adjustment software *Australis 6*. 27 images of our control point field from different viewpoints are used to achieve a reliable camera calibration (see Figure 2). It can be assumed that the image axes are perfectly orthogonal and the image aspect ratio is 1.0. The radial distortion with a maximum displacement of 1.8 pixels is eliminated by resampling all images using bicubic interpolation.

We prefer an algebraic representation of the orientation data instead of the physical model to simplify the transformation from sensor coordinates to image coordinates [4]. The imaging geometry is modeled in terms of line-preserving 3×4 *projection matrices* **P** and a constant *calibration matrix* **K** with a principle distance in pixels:

$$\mathbf{K} = \begin{bmatrix} 2909.1 & 0 & 748.4 \\ 0 & 2909.1 & 408.7 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

P is a composite matrix computed from the calibration matrix **K**, *rotation matrix* **R** and the position of the *projection center* **C**

$$\mathbf{P} = \mathbf{KR}\left[\mathbf{I}\,\middle|\,\mathbf{-C}\right], \tag{2}$$

where **I** denotes the identity matrix. The image coordinates $\mathbf{x} = (x, y)^{\mathsf{T}}$ of an object point $\mathbf{X} = (X, Y, Z)^{\mathsf{T}}$ is computed as follows:

$$w \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3}$$

Figure 3: *Sharp XG-P10XE* LCD-projector and *Liesegang 3600 AV* slide-projector

2.2.2 Texture Projector

A *Sharp XG-P10XE* LCD-projector with a resolution of 1024×768 pixel is used to map various textures onto the object (see Figure 3). Calibration of the LCD-projector is also necessary for some of the surface reconstruction algorithms such as shape-from-texture (see Section 3.4). The interior orientation is obtained from the technical reference sheet of the LCD-projector and transformed into pixel metric. The exterior orientation, however, must be computed manually. Again, we used the *Australis* software to compute the exterior orientation by modeling the LCD-projector as a special camera. The estimated orientation is verified by reprojecting a synthetic chessboard structure for which the 3D coordinates are already known.

2.2.3 Light Projector

The LCD-projector produces raster artifacts with blank texture images. So, an additional slide-projector is required to obtain images suitable for photometric stereo. In our setup we use a *Liesegang 3600AV* slide-projector to model a point light source (see Figure 3). Slide-projector positions are measured manually and the illumination directions are verified by analyzing the highlights on five spheres, which are placed in the scene for the point cloud registration of the laser scanner (see Figure 4). Using this simple method we obtained an accuracy of 2.5 degrees.

2.3. Image Acquisition

The digital camera is firmly mounted on a *KUKA KR C1* industrial robot arm (see Figure 2). Due to various safety reasons, we have adjusted the camera and captured the images remotely through a firewire connection between a notebook and the camera.
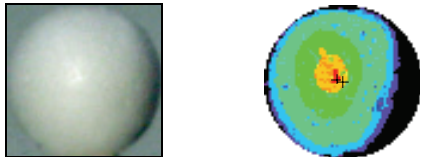

Figure 4: Styrofoam sphere with the estimated highlight position to verify the light direction

Figure 5: Orientation of the reference image (red), narrow- and wide-baseline stereo images (green) and the linear motion sequence (yellow)

This also prevents the possible vibrations during image acquisition and improves the image quality. The robot enables a stable and repeatable camera motion with a relative precision of 0.1 mm. Images are captured at 13 different positions, which have a horizontal baseline of approximately 1 m and a vertical baseline of 0.5 m. (see Figure 5).

A computer-controlled *MICOS Pollux* high resolution positioning device (turntable) is configured to simulate a camera motion around the scene. Rotating the turntable in 9-degrees-steps, we acquired 40 images of the object (see Figure 6). It was not possible to rotate the LCD-projector and the slide-projector synchronous with the turntable, so we used the ambient light in the room and projected no texture on the object for the camera ring images.

2.4. Ground Truth Generation

A *Mensi S25* short range triangulation laser scanner [22] is used to generate the independent ground truth information (see Figure 7). The scanner is specified to have depth accuracy up to 0.2 mm for our object distance of 4 m. The ground truth is acquired from 7 viewing directions in order to generate a 3D point cloud for the entire scene. Having registered the individual views, we obtain an overall accuracy of 0.8 mm.
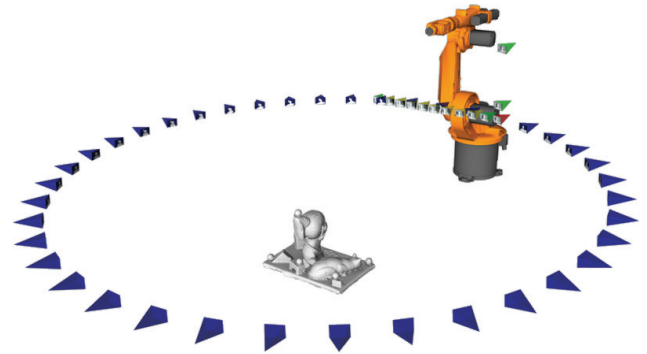

Figure 6: Overview of the imaging geometry. The ring (blue) is simulated using a computer controlled turntable
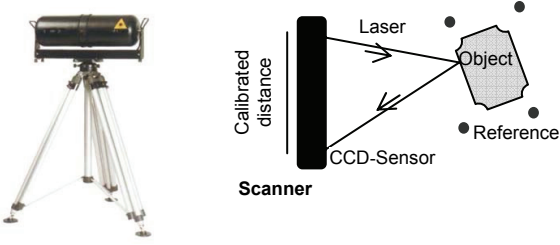
Figure 7: Triangulation principle of *Mensi S25* laser scanner

This meets the requirements, because the depth resolution from images (e.g. pixel-accurate wide-baseline matching in Section 3.1.3) is expected to be 1.5 mm. Background and noisy points are manually removed. The registered and segmented point cloud has a density of approximately 2 mm (see Figure 8).
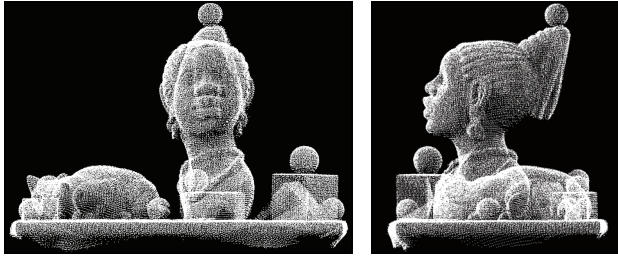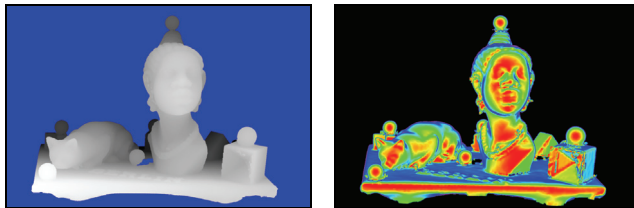

Figure 8: Laser scanner point cloud of the ground truth

Based on the scanned 3D data, we derived two ground truth data maps for the reference image: a normalized depth map, where the gray value of each pixel is related to a depth value of the scene, and a color-coded surface orientation map, where the color of each pixel is related to an orientation vector (see Figure 9).

a.) Normalized depth map   b.) Surface orientation map

Figure 9: Ground truth data for the reference image

3. Surface Reconstruction from Images

The basic idea of all image-based surface reconstruction techniques is that the color of light coming from a particular point in a scene can be rewritten as a function of the environment parameters, such as settings and the positioning of camera, lighting conditions, reflectance properties of surface and background.

Image-based surface reconstruction is a challenging computer vision problem under general imaging conditions. However, controlling the imaging environments simplifies this task, and makes surface reconstruction possible. In the following sections, we explain briefly the investigated surface reconstruction methods.

3.1. Shape from Stereo

Stereo is a well-known technique and works analogue to human vision. The scene must be observed from different viewpoints and for corresponding homologous image points the 3D object point is computed using the orientation information. This process is called triangulation. The correspondence problem is the most challenging part of this algorithm, especially for untextured and occluded areas.

Various global optimization strategies, e.g. dynamic programming [12], graph cuts [10], belief propagation [8] or semi-global matching [6], are used to achieve state-of-the-art results. Standardized test data for binocular stereo is already available [16, 1] including a comfortable online evaluation of the results.

3.1.1   Trinocular Stereo

A binocular image match can be verified using a third image. Given two corresponding points the position in the third image can be predicted. The similarity of this point may support or reject the match. The convergent stereo triplets of our benchmarking dataset are rectified, so that they correspond to the stereo normal case [5].

The important advantage of this method is that the correspondence analysis is simplified to a linear search along one of the image axes (see Figure 10 and 11). An additional property is that the differences (disparities) between horizontal and vertical corresponding points are equal.

3.1.2   Narrow-baseline Stereo

A smaller distance between the cameras leads to very similar images and simplifies the correspondence analysis (see Figure 10). However, as the triangulation angle gets smaller, accurate spatial intersection gets harder. This phenomenon is also called glancing intersection.

3.1.3   Wide-baseline Stereo

A larger distance between the cameras makes the spatial intersection more accurate. But, in this case, perspective distortions and stronger occlusions make the correspondence analysis more difficult (see Figure 11).
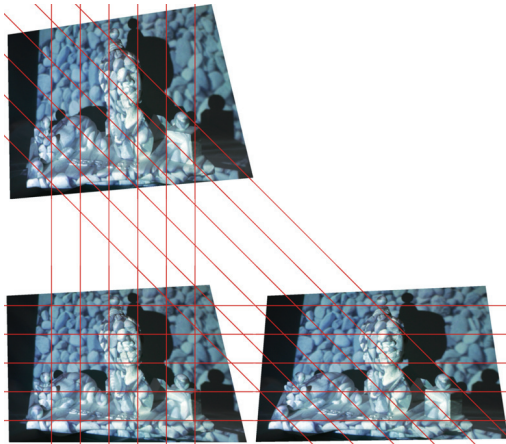
Figure 10: Example of a rectified narrow-baseline triplet with overlaid epipolar lines (23 textured image triplets)
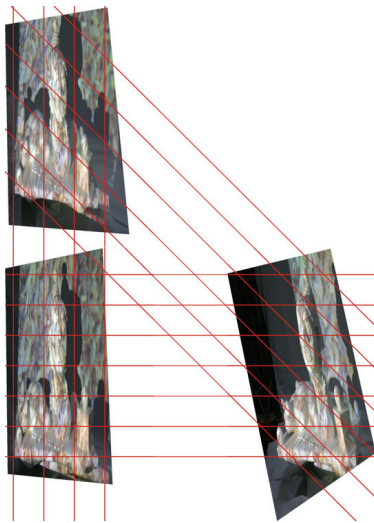

Figure 11: Example of a rectified wide-baseline triplet (23 image textured triplets)

3.2. Shape from Motion

Multi-view stereo combines the advantages of narrow- and wide-baseline stereo. Using neighboring images, the correspondence analysis is simplified, and the baseline for a spatial intersection is extended by using feature tracking (see Figure 12). In [18] standardized test data for multi-view stereo with an online evaluation was made available.

Figure 12: Examples of the image sequence acquired with a linear camera motion (253 textured images)

3.3. Shape from Silhouette

3.3.1 Space Carving

Shape-from-silhouette is a surface reconstruction method which constructs a 3D shape estimate of an object using multiple contour images of the object (see Figure 13). The output is known as the visual hull. This technique still suffers from concavities and insufficient views.

A simple approach is the volumetric reconstruction using voxel space carving [13]. The final reconstruction is still coarse. A more accurate representation is possible with marching intersections [20] or polygonal reconstructions using generalized cone intersections [14].

Figure 13: Examples of the turntable sequence with segmented contour images (40 images)

3.3.2 Voxel Coloring / Photoconsistency

The analysis of consistent scene colors and texture can be used to refine the visual hull [19]. Assuming Lambertian reflection and textured surfaces a reconstruction of concave areas is possible.

3.3.3 Shadow Carving

This approach analyzes self-shadows on the surface which may indicate concavities [15]. Here, the illumination direction has to be given and the detection as well as the categorization of shadows is difficult.

3.4. Shape from Texture

The basic principle behind shape from texture is the perspective distortion of the individual texture elements (texel). Their variation across the image allows the estimation of the surface shape. Structured light is one of the commonly used methods [17]. It is based on active triangulation by replacing the second stereo camera with a calibrated LCD-projector.

Multiple stripes are projected onto the scene, in order to distinguish between stripes that are coded either with different brightness or different colors (see Figure 14 and 15). Therefore, the correspondence problem is solved. Depth information can be computed out of the distortion along the detected profiles.
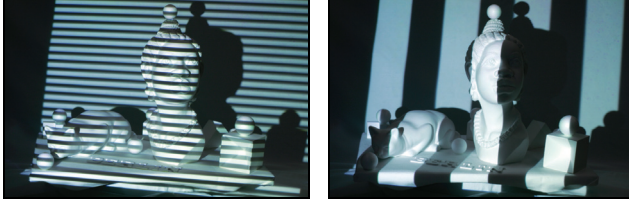
Figure 14: Examples of binary coded textures (9 images with horizontal and 10 images with vertical stripes)
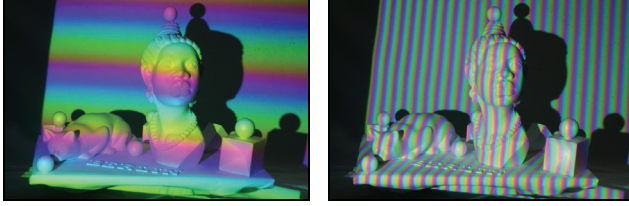

Figure 15: Examples of color spectrum textures (9 images with horizontal and 10 images with vertical pattern)

We projected a standard binary pattern and a rainbow like color spectrum onto the scene. Every stripe of length n in the color spectrum S_{RGB} is generated using the below equations, where i denotes the raster position in the spectrum image and G_{MAX} denotes the maximum intensity value in every color channel [11].

$$S_R = \sin\left(\frac{i}{n} \cdot \pi\right) \cdot \left(\frac{G_{MAX}}{2} - 1\right) + \frac{G_{MAX}}{2}$$

$$S_G = \sin\left(\left(\frac{2}{3} + \frac{i}{n}\right) \cdot \pi\right) \cdot \left(\frac{G_{MAX}}{2} - 1\right) + \frac{G_{MAX}}{2}, i = \{0,...,n\} \qquad (4)$$

$$S_B = \sin\left(\left(\frac{4}{3} + \frac{i}{n}\right) \cdot \pi\right) \cdot \left(\frac{G_{MAX}}{2} - 1\right) + \frac{G_{MAX}}{2}$$

3.5. Shape from Shading

Shape-from-shading [7] is a single image technique to estimate a 3D surface with Lambertian reflectance properties from a known illumination direction. Recent techniques can deal with more complex reflectance functions, i.e. Helmholtz reciprocity [24], light transport constancy [25] or appearance clustering [26].

Photometric stereo [9] is an extension that uses at least three monoscopic images with different illumination directions to achieve a unique result (see Figure 16). These reflectance based methods determine surface orientations instead of depth information.

a.) Left   b.) Top   c.) Right

Figure 16: Reference image under different illumination configurations (3 Lambertian-like images)

3.6. Shape from (De-)Focus

The last approach uses monoscopic image sequences with varying focal length [2]. A segmentation of sharp and blurred image parts provides information about the actual focused depth (see Figure 17). This method requires special sensor properties with an extreme small depth of field (e.g. microscope).
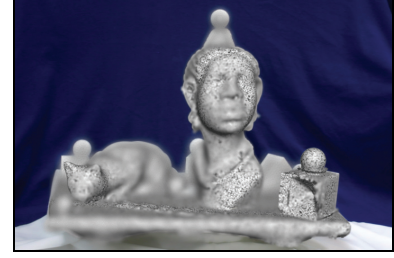

Figure 17: Synthetically rendered image using depth of field

4. Evaluation

The surface reconstruction methods, which are investigated in this study, have various output formats such as disparity maps, depth maps, volumetric data or surface orientation maps (see Table 1). In order to compare these different outcomes quantitatively, one should first convert them to a comparable form.

Method	Output
Stereo	Disparity map, depth map
Motion	Polyhedron, point cloud
Silhouette	Volumetric data, polyhedron
Texture	Disparity map, depth map
Shading	Surface orientation, needle map
Focus	Depth map

Table 1: Output of investigated methods

Fortunately, except for the surface orientation map, all other formats can easily be converted to a depth map, if image orientation is available. As we already provide the image orientation, quantitative comparison can be performed through depth maps. For the evaluation of the reflectance based methods surface orientation maps are preferred.

Once the reconstruction outcome is converted into a depth/orientation map, it is possible to compare it with the ground truth depth/orientation map. This comparison is done according to two criteria. The first criterion is accuracy. Accuracy is a measure of closeness of the values of the reconstruction and the ground truth. Second criterion is completeness. It is a measure of overall success of the algorithm. The more the scene is modeled, the more complete is the reconstruction.

Let us denote the ground truth map with G and the reconstructed map with R. The accuracy of the reconstruction is computed using the difference between these two maps as shown in Equation 5. In this equation M is the set of pixels used in evaluation, n is the total number of pixels in M, and g is the normalization factor. Normalization factor is either the maximum depth value (i.e. $g = 255$) or the maximum norm of the difference vector given in the equation (i.e. $g = 2$).

$$a_{R,M} = 1 - \sqrt{\frac{1}{n} \sum_{(i,j) \in M} \left| \frac{R(i,j) - G(i,j)}{g} \right|^2} \qquad (5)$$

The accuracy term can be used to rank different reconstruction results. However, accuracy alone would not be sufficient for a fair ranking. If there are pixels on the map, whose depth/orientation values cannot be correctly reconstructed, this should also be taken into consideration. Completeness is defined as the ratio of well-reconstructed pixels to all pixels as given in Equation 6. δ is a threshold for error tolerance.

$$c_{R,M} = \frac{1}{n} \sum_{(i,j) \in M} \left(\frac{R(i,j) - G(i,j)}{g} \le \delta \right) \qquad (6)$$

Using the formulas provided above it is possible to measure the accuracy and the completeness of a reconstruction at specific regions. This allows us to investigate the success of each method in reconstructing challenging cases, such as self occlusion, concavities, sharp surface discontinuities etc. Thus, mostly inspired by [16], we provide masks (see Figure 18), which can be used to compute the accuracy and completeness according to the following criteria.

a.) Object mask   b.) Object without shadows

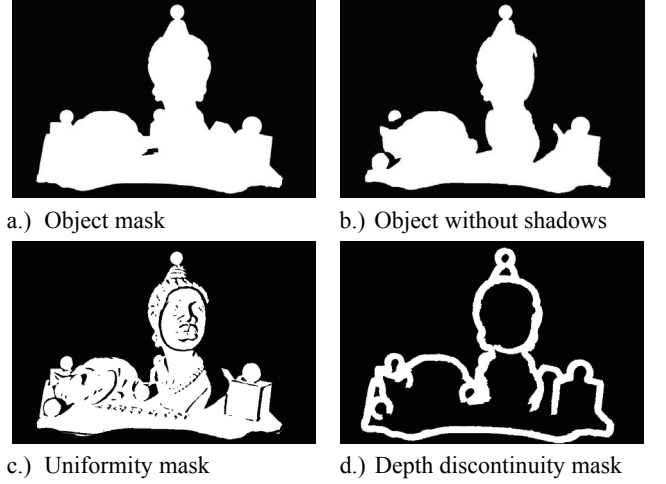c.) Uniformity mask   d.) Depth discontinuity mask

Figure 18: Masks for investigating detailed aspects

- **Uniformity:** These are the regions where the deviation of the image gradients is less than a given threshold.
- **Occlusion:** These are the regions that are occluded in the matching image.
- **Depth discontinuity:** These are the regions where the deviation of depth values of the pixels is larger than a given threshold.

5. Conclusions

In this study we introduce a true benchmark dataset for performance evaluation of shape-from-X algorithms and a test procedure for evaluating the reconstruction results. Our aim in this work is to support an objective comparison of different approaches and to provide an informative basis for the combination of reconstruction methods. Researchers are invited to download this benchmarking dataset from our web server [21] and return their results, in order to describe and improve the performance of automatic surface reconstruction techniques.

6. Acknowledgements

References

[1] P. Courtney, N. Thacker and A. Clark: *Algorithmic modeling for performance evaluation*, ECCV Workshop on Performance Characteristics of Vision Algorithms, Cambridge, UK, 1996.

[2] P. Favaro and S. Soatto: *A geometric approach to shape from defocus*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 406-417, 2005.

[3] W. Förstner: *10 Pros and cons against performance characterization of vision algorithms*, ECCV Workshop on Performance Characteristics of Vision Algorithms, Cambridge, UK, 1996.

[4] R.I. Hartley and A. Zisserman: *Multiple view geometry in computer vision*, Cambridge University Press, 607 p., 2000.

[5] M. Heinrichs and V. Rodehorst: *Trinocular rectification for various camera setups*, Symp. of ISPRS Commission III - Photogrammetric Computer Vision PCV'06, Bonn, Germany, pp. 43-48, 2006.

[6] H. Hirschmüller: *Stereo vision in structured environments by consistent semi-global matching*, IEEE Conf. on Computer Vision and Pattern Recognition CVPR'06, Vol. 2, New York, NY, USA, pp. 2386-2393, 2006.

[7] B.K.P. Horn and M.J. Brooks: *Shape from shading*, MIT Press, Cambridge, MA, 1989.

[8] A. Klaus, M. Sormann and K. Karner: *Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure*, IEEE Int. Conf. on Pattern Recognition ICPR'06, Hong Kong, China, pp. 15-18, 2006.

[9] R. Klette, R. Kozera and K. Schlüns: *Reflectance-based shape recovery*, Handbook of Computer Vision, vol. 2, Academic Press, pp. 531-590, 1999.

[10] V. Kolmogorov and R. Zabih: *Computing visual correspondence with occlusions via graph cuts*, IEEE Int. Conf. on Computer Vision ICCV'01, vol. 2, Vancouver, Canada, pp. 508-515, 2001.

[11] A. Koschan and V. Rodehorst: *Dense depth maps by active color illumination and image pPyramids*, In: F. Solina, W.G. Kropatsch, R. Klette, R. Bajcsy (Eds.), Advances in Computer Vision, Springer-Verlag, Vienna, Austria, pp. 137-148, 1997.

[12] C. Lei, J. Selzer and Y.H. Yang: *Region-tree based stereo using dynamic programming optimization*, IEEE Conf. on Computer Vision and Pattern Recognition CVPR'06, New York, NY, USA, pp. 2378-2385, 2006.

[13] W. N. Martin and J. K. Aggarwal: *Volumetric description of objects from multiple views*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 5, no. 2, pp. 150-158, 1983.

[14] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler and L. McMillan: *Image based visual hulls*, ACM Computer Graphics (SIGGRAPH'00), pp. 368-374, 2000.

[15] S. Savarese, H. Rushmeier, F. Bernardini and P. Perona: *Shadow carving*, IEEE Int. Conf. on Computer Vision ICCV'01, Vancouver, CA, 2001.

[16] D. Scharstein and R. Szeliski: *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*, Int. Journal of Computer Vision, vol. 47, no. 1-3, pp. 7-42, 2002.

[17] D. Scharstein and R. Szeliski: *High-accuracy stereo depth maps using structured light*, IEEE Conf. on Computer Vision and Pattern Recognition CVPR'03, vol. 1, Madison, WI, USA, pp. 195-202, 2003.

[18] S. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski: *A comparison and evaluation of multi-view stereo reconstruction algorithms*, IEEE Conf. on Computer Vision and Pattern Recognition CVPR'06, vol. 1, New York, NY, USA, pp. 519-526, 2006.

[19] S. Seitz and C. Dyer: *Photorealistic scene reconstruction by voxel coloring*, Int. Journal of Computer Vision, vol. 35, no. 2, pp. 151-173, 1999.

[20] M. Tarini, M. Callieri, C. Montani, C. Rocchini, K. Olsson and T. Persson: *Marching intersections: an efficient approach to shape-from-silhouette*, IEEE Workshop on Vision, Modeling and Visualization (VMV), Erlangen, Germany, pp. 283-290, 2002.

[21] ISPRS Photogrammetric Computer Vision and Image Analysis Working Group III/2 - *Surface reconstruction*, http://www.cv.tu-berlin.de/isprs/wgiii2/, 2004 – 2008.

[22] W. Böhler and A. Marbs, *3D scanning instruments*, Int. Workshop on Scanning for Cultural Heritage Recording, Corfu, Greece, 2002.

[23] R. Klette, A. Koschan, K. Schlüns and V. Rodehorst: *Evaluation of surface reconstruction methods*, New Zealand Image and Vision Computing Workshop, Lincoln, Canterbury, pp. 3-12, 1995.

[24] T.E. Zickler, P.N. Belhumeur and D.J. Kriegman: *Helmholtz stereopsis: exploiting reciprocity for surface reconstruction*, Int. Journal of Computer Vision, vol. 49, no. 2-3, pp. 215-227, 2002.

[25] J.E. Davis, R. Yang and L. Wang: *BRDF invariant stereo using light transport constancy*, IEEE Int. Conf. on Computer Vision ICCV'05, vol. 1, pp. 436-443, 2005.

[26] S.J. Koppal and S.G. Narasimhan: *Clustering appearance for scene analysis*, IEEE Conf. on Computer Vision and Pattern Recognition CVPR'06, vol. 2, pp. 1323-1330, 2006.

[27] R. Zhang, P.S. Tsai, J.E. Cryer and M. Shah: *Shape-from-shading: a survey*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 8, pp. 690-706, 1999.