Evaluation of Surface Reconstruction Methods

Reinhard Klette, Andreas Koschan, Karsten Schlüns and Volker Rodehorst

Computer Vision Group, FR 3-11
Computer Science Department, Berlin Technical University, D-10587 Berlin, Germany
http://www.cs.tu-berlin.de/~cvworld
{klette,koschan,karsten,vr}@cs.tu-berlin.de

Abstract

In this paper we discuss the evaluation problem of surface reconstruction algorithms following the methodologies of static, dynamic and photometric stereo analysis. It is important to answer such questions as 1) For what kind of surfaces and 3-D objects an algorithm behaves either well or bad? 2) How accurate are the reconstruction results of an algorithm under specified circumstances? What measure can be used to evaluate reconstruction accuracy? 3) How to compare reconstruction results following different methodologies? 4) What algorithm can be suggested for a specific application project? and so on. So far we present some proposals and first quantitative and qualitative results for answering such questions. In our opinion a methodology for evaluating surface reconstruction algorithms is still at its beginning.

1. Introduction

Reconstructed surface patches are approximations of 3-D object faces. They can include absolute or relative depth information. *Isolated measurements* of 3-D point positions are approximate locations of specific surface points. Reconstructed surface patches and isolated measurements characterise the two typical aims of surface reconstruction methodologies in computer vision as static, dynamic or photometric stereo analysis.

An *experiment* for the performance evaluation of stereo analysis results is described by specifying its input data and the available ground truth, by defining an error criterion, by specifying the selected algorithms which are used, and by a discussion of the experimental results and the conclusions. Two scene spaces PACKAGE and BEETHOVEN are used as input image sources, cp. Fig. 1.

Quantitative evaluations of algorithms are still a rare case in the computer vision literature. Accuracy is a relation between *ground truth* and calculated result. For surface reconstruction algorithms, the geometrical locations of 3-D scene object surfaces should act as ground truth, and the reconstructed surface patches or isolated measurements are the calculated results. If synthetic objects would be used then such a ground truth is known. For some real objects as the box in PACKAGE scenes, the object sizes can be measured, cp. Tab. 3. A quantitative evaluation requires an *error criterion function*, e.g. the average of all deviations in reconstructed sizes of the box in PACKAGE scenes is such a function. If such functions are not available then the qualitative visual comparison of reconstructed surfaces with the given 3-D objects is used as a qualitative error criterion. The used representation method may influence this evaluation. In this paper, a qualitative evaluation is demonstrated for different reconstruction algorithms in the second scene space BEETHOVEN.

In this second scene space the ground truth is not available. However, even if the geometrical surface data would be available then still a proper confidence measure has to be selected: How to measure geometrical differences between two 3-D surfaces representing partial views of the same 3-D object? The complexity of this problem increases if only rela-

tive depth values can be reconstructed, i.e. the reconstructed surface patches must be scaled with some scaling factor.

Figure 1: Input pictures of the scene spaces PACKAGE (box, bottle etc.) and BEETHOVEN (plaster statue, alone or in a desk environment).

For image acquisition a 3CCD-Colour-Camera DXC730P from Sony was used. Gain control and gamma correction were turned off. Automatic white balancing was used. The digitisation process was done by a DATACELL s2200-frame grabber in a Sun IPX workstation. The focal length used with static stereo (Section 2) and dynamic stereo (Section 3) was 23 mm. The focal length used with the shading based approaches (Section 4) was ca. 60 mm. The scenes were illuminated by one or several slide projectors (dynamic stereo and shading based approaches) or fluorescent ceiling lamps (static stereo).

The following coordinate systems are used to model the relations between scene space objects and projected images:

(X_w, Y_w, Z_w) denote the 3-D coordinates of object surface points **P** in the *world coordinate system* ,

(X_c, Y_c, Z_c) denote the 3-D coordinates of **P** in the *camera coordinate system* ,

f is the distance between the image plane and the projection centre *(focal length)*,

(x_u, y_u) are *non-distorted* image coordinates of (X_c, Y_c, Z_c) assuming an ideal pinhole camera,

(x_d, y_d) are *distorted* image coordinates, differing from (x_u, y_u) by radial lens distortion, and

(x_f, y_f) are *device-dependent* coordinates of (x_d, y_d) in the digitised image.

The Z-axis Z_c of the camera coordinate system coincides with the optical axis, and it is pointing into the scene space. If only relations between camera coordinates and non-distorted image coordinates are discussed then world and camera coordinates are assumed to be identical. For simplification, (x, y) and (X, Y, Z) are used in this case, without indices u , w or c. All coordinates and parameters are measured at the same scale, e.g. μm. The only exception are discrete coordinates (x_f, y_f) for the digitised image which are given in (sub-) pixels. Central projections are assumed in Sections 2 and 3, and parallel projections in Section 4.

For camera calibration, internal camera parameters as well as geometric relations between camera coordinates and world coordinates have to be pre-calculated, and these data have essential influence on the accuracy of the surface reconstruction results. For example, for describing the affine transform from world coordinates (X_w, Y_w, Z_w) into camera coordinates (X_c, Y_c, Z_c), a rotation matrix **R** and a translation vector **T** have to be calibrated. The method [1] was used.

For solving the *evaluation of camera calibration method* problem, an extension of this method was necessary, see [2]. Using this extension of Tsai's method, the accuracy of calibration results was evaluated. Different real and synthetic calibration objects (planes with

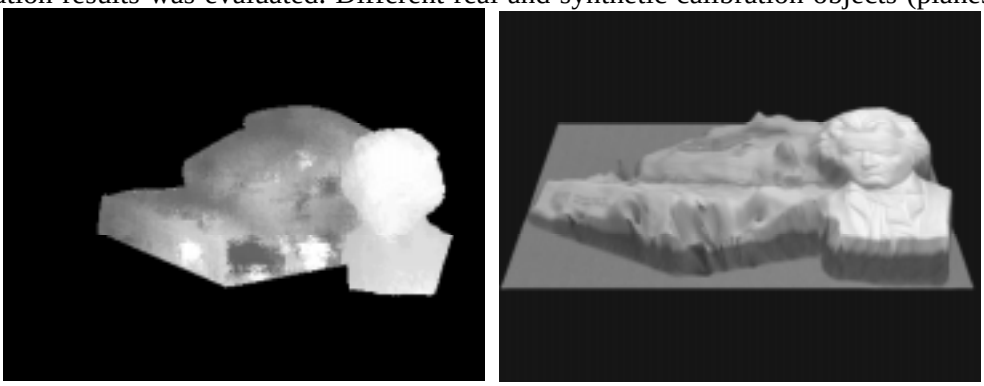

Figure 2: Gray coded dense depth map obtained with chromatic block matching (left) and texture mapping for these depth values (right).

calibration points in the latter case) were used for selecting a simple, but sufficient calibration object. Using synthetic objects it was possible to prove statistically that near-optimum accuracy is achievable already by three calibration planes, each with about 20 calibration points (e.g. an "open cube").

2. Static Stereo Analysis

The key problem in static stereo analysis is how to find the corresponding points in the left and in the right image, referred to as the *correspondence problem*. Whenever the corresponding points are found, the depth can be computed by triangulation. Stereo techniques can be distinguished by either matching edges and producing sparse depth maps, or by matching all pixels in the entire images and producing dense depth maps.

An evaluation of static stereo approaches was performed for selecting methods for further study [3]. For such an evaluation it is not suitable to evaluate a selected single processing step (e.g. a stereo matching algorithm) without taking into consideration the interdependence of all processing steps. For example, matching is simple if the correspondence search is reduced to characteristic features like junctions in the image. Although, the matching problem can be solved perfectly for those features. However, the obtained results are not suitable for a dense surface reconstruction process. The main emphasis of the investigations presented in [3] was on stereo matching because this processing step has a great influence on the quality of the computed results. In addition, the entire correlation and the mutual dependence of the processing steps within a stereo system were taken into consideration. The experiment can be described as follows:

Input data and ground truth: Correspondence methods are evaluated with regard to their suitability within a stereo system for the recovery of the surface geometry of 3-D objects in near distance to the cameras (approximately 1 m). All methods are applied to a series of indoor images. Gray value and colour images are used as possible input data.

Error criterion: Qualitative evaluations of calculated correspondences based on visualisations as shown in Figs. 2 and 3.

Algorithms: After investigating 73 different stereo methods, eight selected and two new matching techniques were implemented.

Results and conclusions: As a result of these comparisons an edge-based approach using disparity histograms [4] and a technique based on block matching [5] were selected for further study. The quality of the matching results improves for both techniques if colour information is used instead of grey value information [6]. As a result the $I_1 I_2 I_3$ colour space suggested by [7] provides the best information for stereo matching. The selection of the colour measure had no significant influence on the results in this investigation. The

precision of the matching results always improved by 20 to 25% when using colour information instead of grey value information.

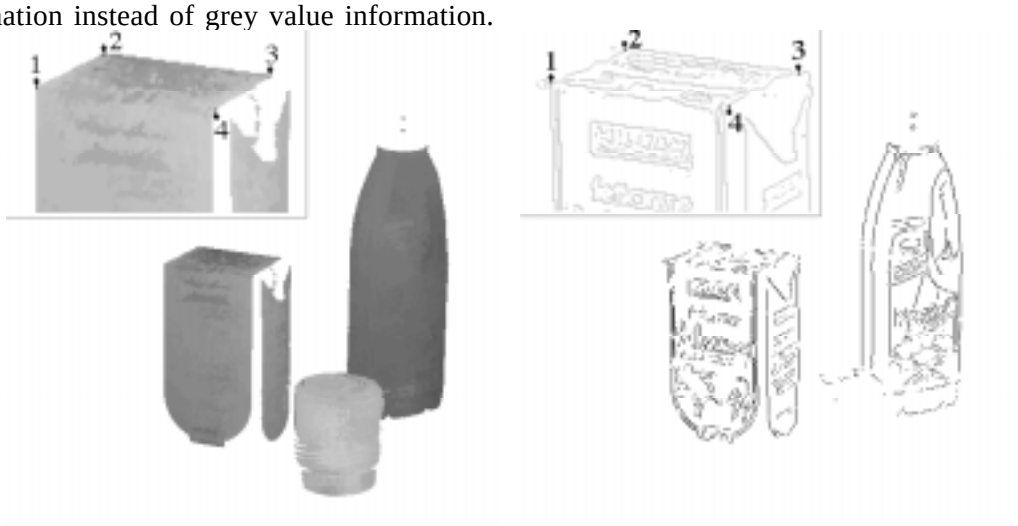

Figure 3: Depth values obtained to an PACKAGE stereo image when applying chromatic Block Matching (left), and points where the edge-based approach has matched zero-crossings (right). The camera was about 1.80 m away from these objects, and the baseline distance between both cameras was 11.7 cm.

Dense depth maps obtained with the chromatic block matching method are shown in Figs. 2 and 3 at the left. The edge-based approach allows computations of corresponding points only at edges, see Fig. 3 at the right. For edge extraction the Marr-Hildreth operator was used at three resolutions (σ_1 = 1.41, σ_2 = 3.18, and σ_3 = 6.01). Zero-crossings in the LoG filtered images constitute the features in the succeeding matching process.

3. Motion Analysis

Shape from motion is often cited as one of the basic computer vision approaches. Rigid objects are projected into the image plane assuming a certain camera model. For a time sequence of such projections, motion vectors have to be computed. Based on these vectors, certain shape values of the objects can be determined leading to a (partial) 3-D representation of the projected objects.

The computation of high-accuracy dense motion vector fields is the most critical issue of surface from motion, cp. [8, 9] for evaluations of optical flow algorithms. Approaches for solving this task can be classified as point-based differential methods, as region based matching methods, as contour-based methods, or as energy-based methods.

Point-based differential methods use spatial and time gradients of the image irradiance function. The point-based differential methods are favourite candidates for surface reconstruction approaches since complete and dense motion fields can be computed. Several motion detection algorithms were evaluated [2, 9] using the source code of [8] if offered by this source. The Anandan method, cp. [8], was selected for further study (note that *dense* flow fields are required for shape from motion).

Shape from motion results were evaluated for objects placed on a rotating disc. The radius of the disc is known, and the camera calibration method is also used for calculating the rotation axis in camera coordinates. For this case of rotational motion [2], for correspondence calculation the epipolar constraint of static stereo can be adopted.

Dynamic stereo analysis for such a rotating disc allows depth computation for corresponding points in consecutive images. Assume that during image acquisition of an object placed on the rotating disc, projections of a visible surface point $\mathbf{W}$ (in the world coordinate system) are given as points $\mathbf{C}_1$ and $\mathbf{C}_2$ in the image plane of the camera-centred coordinate system at consecutive time slots t and $t+1$. The task consists in calculating the coordinates of $\mathbf{W}$ based on the known pixel positions of points $\mathbf{C}_1$ and $\mathbf{C}_2$, and based on the

calibration results, e.g. rotation matrix **R** and translation vector **T**, cp. Section 1. Solutions are given in [2].

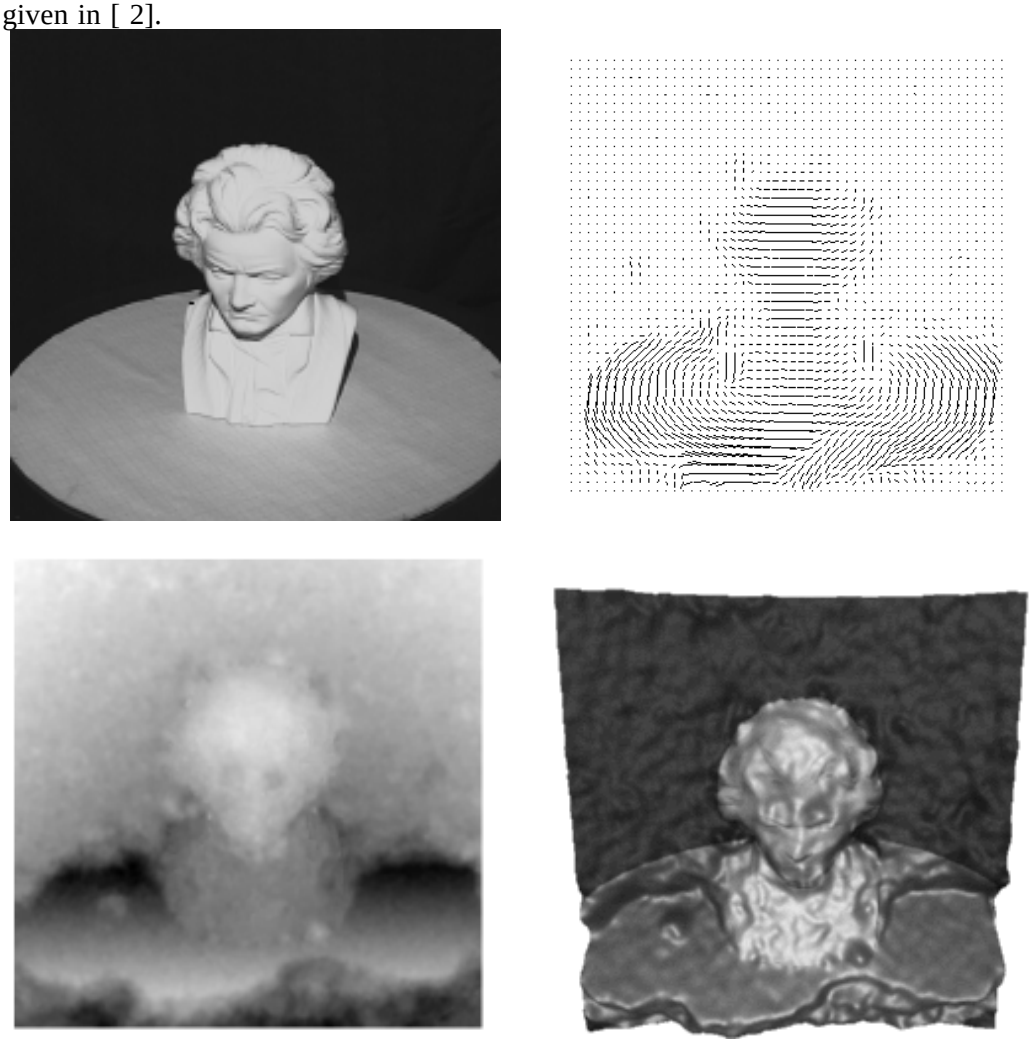

Figure 4: The plaster statue on the rotating disc, a needle representation of the optical flow field calculated by the Anandan method, cp. [8], and two representations of the reconstructed 3-D surface: depth map (below left) and texture mapping (below right).

Unfortunately, automatic dense flow vector field computation is not available at a quality level allowing good quality reconstructions of real object surfaces. By adding noise to ideal motion vector fields of synthetic objects it can be proofed that very small disturbances already have a great impact on the reconstructed surfaces. In Fig. 4 it is illustrated what quality of reconstruction could be achieved in case of a BEETHOVEN scene.

Dense optical flow field computation does not lead to complete surface reconstructions at acceptable quality. However, singular correspondences ("sparse flow fields") can be calculated for real objects with relatively good accuracy, e.g. by edge-based correspondence analysis, allowing computations of 3-D positions of surface points for the calculated feature points. Matched point positions of a PACKAGE scene are illustrated in Fig. 5. A sequence of four input images was used. The rotation angle between two images was 5°. The camera was about 1.80 m away from these objects. The rotational disc is shown in the image just for illustration purposes. The Marr-Hildreth operator was used for edge detection. The known rotation angle is used within the reconstruction procedure.

4. Shading based shape recovery

Shading based methods for surface reconstruction consider the imaging sensor as a photometric measuring device. Assuming special reflection characteristics and known illumination configurations, the measured image irradiances reduce possible surface orientations

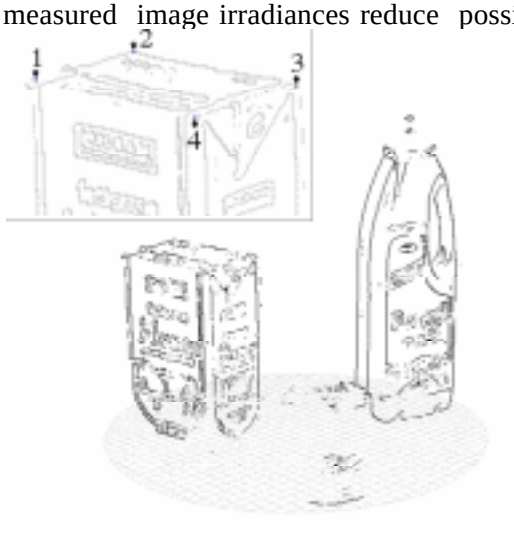

Figure 5: Corresponding points using edge-based motion analysis. Four selected points in corner positions are illustrated.

to a closed curve on the Gaussian sphere. For Lambertian reflection these curves are circles. Several approaches exist to achieve a unique surface in spite of this ambiguity in image irradiance. Most approaches require hard limitations to the object world. This implies that these approaches are theoretically very interesting but not useful in practice.

However shading based methods have the important advantage that they calculate dense three-dimensional surface features. Surface orientation or relative depth can be assigned to each visible surface point in the image, except for sharp discontinuities.

In our work we focus on methods which are independent of the actual surface albedo. Albedo independence allows the surface to be of an arbitrarily coloured texture. Hence, no reflection factor measurements are needed for the originally unknown surface.

Using at least three light sources surface orientations together with albedo values can be measured for each image point independently. This is known as the *photometric stereo method* [10]. Algorithms following this methodology calculate surface gradients or normals from the shading variations in three images, taken of the objects with light sources in different positions. Opposite to the binocular stereo techniques (Section 2), the visual sensor is fixed in the same position and orientation. Opposite to the dynamic stereo techniques (Section 3), the illumination changes during picture sequence acquisition. Therefore no matching or registration strategy is necessary.

From the surface orientations the depth can be calculated using integration techniques. Comparisons with several approaches have shown that a FFT-based method [11] produces the best results. For *polyhedral reconstructions* also the method described in [12] can be used for depth calculations.

The *inaccurate illumination parameter estimation* problem can be solved if the influence of these parameters on reconstruction can be reduced. Traditionally for the photometric stereo method [10], the accuracy of the orientation determination depends strongly on the estimation of the light source parameters, i.e. the light source strength and the illumination direction. On the other hand, table look-up techniques exist which need no know–ledge of these parameters, since the table look-up is built with an calibration object [13]. Table look-up techniques in general have the disadvantage that only objects with surface materials can be analysed which have exactly the same reflection characteristics as the calibration object. This implies many practical problems, since often it is difficult to generate a calibration object with a special surface material, and since the albedo of the analysed object has to be constant. We developed a table look-up technique which is independent of the sur-

face albedo. The light source parameters remain unknown. In comparison to other table look-up techniques, a further assumption is that the sensor has a linear characteristic.

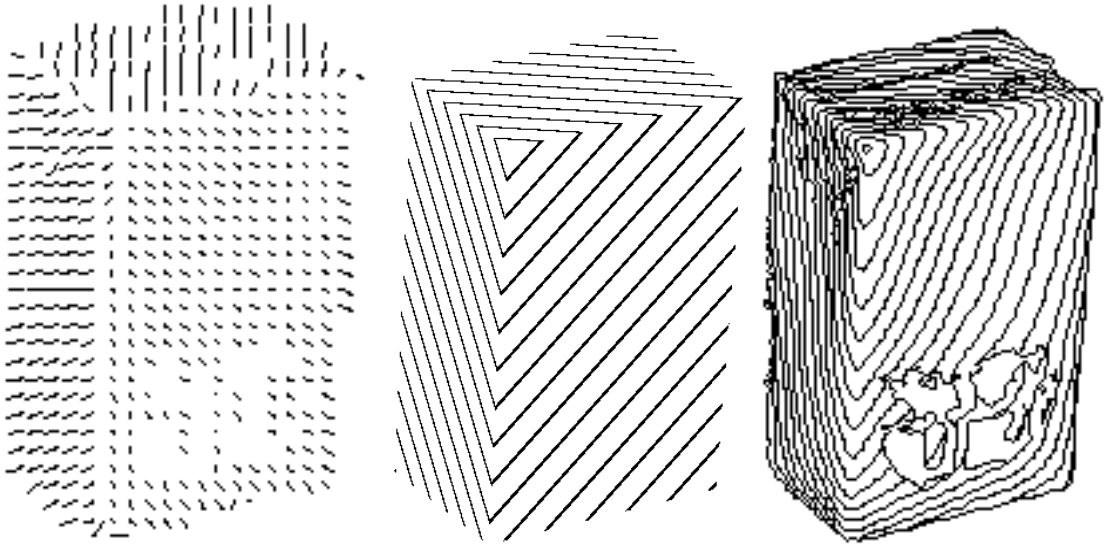

Figure 6: A needle map representation of a reconstructed gradient image using an albedo independent photometric stereo technique (on the left), and iso-depth plots using the approach in [11] (in the middle) and the polyhedron reconstruction method [12] (on the right) for calculating depth from gradients.

face	X	Y	Z
A	0.054525	0.894764	-0.443198
B	-0.836804	-0.244101	-0.490076
C	0.513713	-0.449641	-0.730699

Table 1: Calculated unit normals of three visible box faces.

In Fig. 6 results are shown of photometric stereo analysis of a PACKAGE scene. The reconstructed gradient image (needle map representation on the left) is transformed in depth maps visualised by iso-depth plots. The cow pictured on the box has a very small albedo, i.e. close to zero. Therefore (among other regions) in the cow no surface orientations could be determined.

In Tab. 1 the three determined unit surface normals are listed for the three visible faces in Fig. 6. The angles between these faces are as follows,

$$angle(A,B) = 92.7°,\ angle(A,C) = 92.9°,\ angle(B,C) = 87.8°.$$

The calculated ratios $\mathbf{E}_3$:$\mathbf{E}_1$ = 1.6783, $\mathbf{E}_3$:$\mathbf{E}_2$ = 2.6638, $\mathbf{E}_1$:$\mathbf{E}_2$ = 1.5872 of the edge lengths can be compared to the real edge ratios $\mathbf{E}_3$:$\mathbf{E}_1$ = 1.7579, $\mathbf{E}_3$:$\mathbf{E}_2$ = 2.6508, $\mathbf{E}_1$:$\mathbf{E}_2$ = 1.5079. Using edge $\mathbf{E}_2$ as a reference edge the calculated lengths differ by 5 mm ($\mathbf{E}_1$) and by 1 mm ($\mathbf{E}_3$) from the ground truth.

Traditionally the photometric stereo method restricts the reflection characteristics of the analysed surface to Lambertian reflection. This approximate assumption can be used for many materials. However in general surface reflection is composed of a diffuse and of a specular component. With the specular reflection component more parameters must be known in advance to model this hybrid reflection. These are roughness values and weighting factors, for example. Using colour is a good way to get more information about the scene without introducing more light sources and without making severe restrictions to the surface reflection properties. We apply the dichromatic reflection model. This model can be used to separate the two types of reflection components [14, 15]. Moreover, roughness val-

ues and weighting factors can be extracted from the images. Since the photometric stereo method additionally recovers surface albedo values, illumination independent surface co-

light source	relative strength	estimated p_s	estimated q_s	estimated tilt	estimated slant
1	1.0	-0.387930	0.059682	171.25	21.43
2	0.56271	0.031802	0.556538	86.73	29.14
3	0.52581	0.389363	0.104828	15.07	21.96

Table 2: Light source parameters of the experiment for reconstructing the surface of the Beethoven plaster statue.

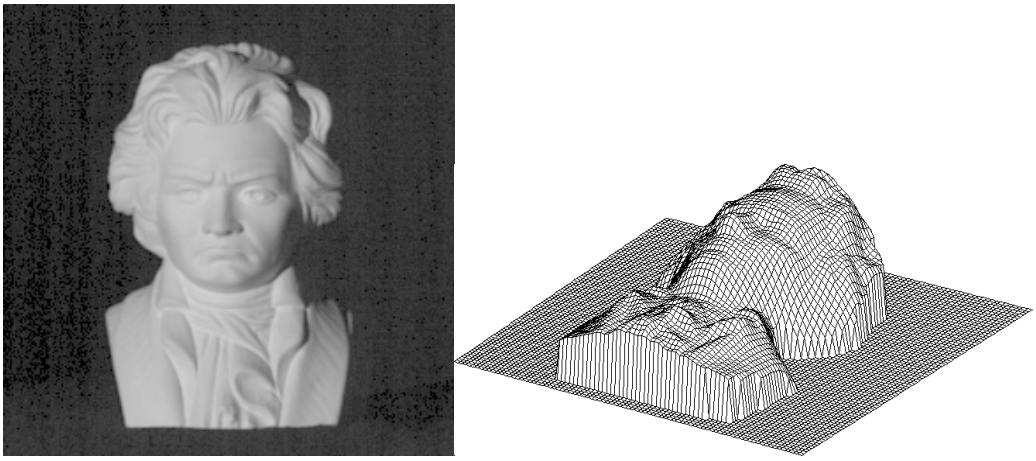

Figure 7: Left: input image of the Beethoven plaster statue illuminated with the first light source, cp. Tab 2. Right: mesh plot of the resultant depth map.

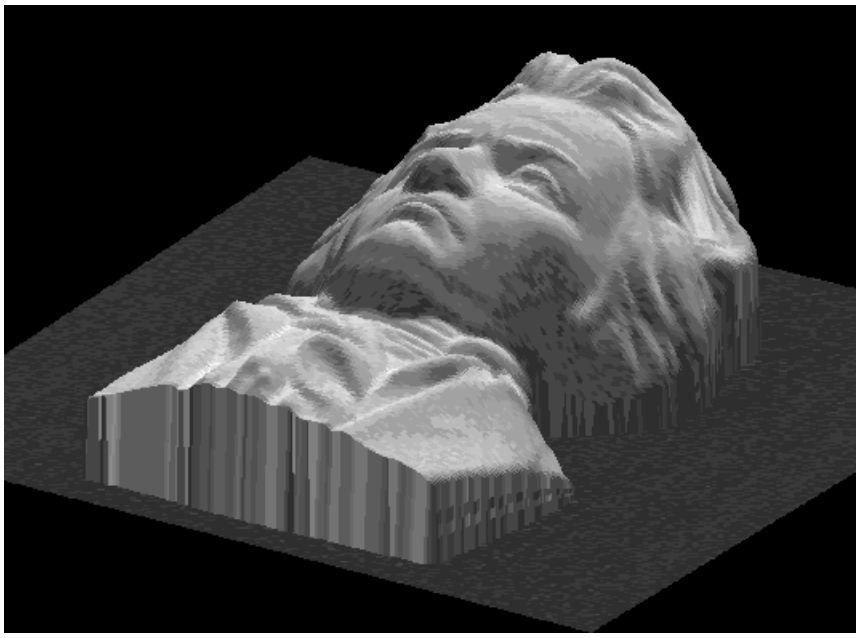

Figure 8: Visualisation of the reconstructed surface of the Beethoven plaster statue using texture mapping.

lour descriptors can be obtained. These colour descriptors allow that a set of CIE tristimulus values can be assigned to each surface element.

The light-source specifications for reconstructing the surface of the Beethoven plaster statue are given in Tab. 2. The FFT-based integration method [11] is used to calculate a relative depth map from the calculated surface orientations.

The albedo of this object is constant. Plaster has approximately a Lambertian reflection characteristic. Therefore no highlight elimination is necessary. The input image illuminated from the first light source is shown in Fig. 7 at the left. Fig. 7 at the right is a visualisation using a mesh plot. The sampling rate of the mesh is three pixels. Therefore details in the depth map are lost.

The albedo independent table look-up method is used to reconstruct the surface orientations in areas where three irradiances are available. In areas illuminated from two light source the surface orientations are projected onto the self-shadow circle of the third source on the Gaussian sphere. In areas illuminated from one source the orientations are projected onto the intersection of the two shadowing sources. In Fig. 8 a surface plot is shown using texture mapping. As texture image the scene illuminated with the second light source was used.

6 Conclusions

Static and dynamic stereo analysis allow to measure absolute depth values. For the box used in the experiments, such measurements are given in Tab. 3. The edge sizes allow a direct comparison with the ground truth. Photometric stereo analysis offers gradient data, see Tab. 1, which can be used for computing a relative depth map. The computed angles were close to 90°, i.e. nearly accurate, cp. Section 4.

distance between		edge length (ground truth)	**static stereo**		**dynamic stereo**
			intensity based	feature based	
$\mathbf{P}_1$	$\mathbf{P}_4$	167.0	169.4	171.6	168.7
$\mathbf{P}_2$	$\mathbf{P}_5$	167.0	166.4	165.5	166.0
$\mathbf{P}_3$	$\mathbf{P}_7$	167.0	167.1	165.2	167.1
$\mathbf{P}_4$	$\mathbf{P}_6$	63.0	56.5	57.1	62.0
$\mathbf{P}_5$	$\mathbf{P}_4$	82.0*	75.4	77.3	88.2
$\mathbf{P}_6$	$\mathbf{P}_7$	82.0*	71.7	78.7	91.3
$\mathbf{P}_7$	$\mathbf{P}_5$	63.0	54.4	57.0	61.6

*If the edge detection algorithm did not find the real positions of the vertices $\mathbf{P}_4$-$\mathbf{P}_7$ (dark box corner points) because their contrast to the background is too low then the edges of the white label were chosen.

Table 3: The comparison of the 3-D point differences (in mm) with the ideal size of the box in the PACKAGE scenes (selected points on the box surface).

Altogether these experiments with the box show that approximate surface data can be obtained with these methodologies based on the used calibration. However, for static and dynamic stereo analysis some corresponding pairs were interactively selected (without optimisation, just with the aim to be close to the object corners).

For reconstructing surface patches the high error rate of static or dynamic stereo analysis results is without doubt, cp. Figs. 2 and 4. The reconstruction of surface patches using static stereo analysis is possible at a very rough level, and dynamic stereo analysis absolutely fails. However, these rough drafts can be of interest for the navigation in 3-D environments, i.e. it is not possible to recognise an object, or to reconstruct the surface in a precise way, but it can be calculated "that there is something" what a robot may view as an obstacle.

The reconstruction of surface patches was possible by photometric stereo analysis in good quality, cp. Figs. 7 and 8. This reconstruction process is also very fast and in on-line only restricted by the time of using three different light source illuminations.

Further approaches exist for reconstructing 3-D surfaces, e.g. structured lightening [16], or shape from shading [13]. They should be included in comparative evaluations of reconstruction results. The technical report [17] is a long version of this conference paper.

Acknowledgments

The first author thanks E.V. Krishnamurthy of the Australian National University, Canberra, for providing good working conditions. Parts of this work (project PARVIS - parallel computation of stereo correspondence algorithms) were funded by the Deutsche Forschungsgemeinschaft (DFG).

References

1. **R.Y. Tsai**, An Efficient and Accurate Camera Calibration Technique for 3d Machine Vision, *Proc. IEEE Conf. Computer Vision And Pattern Rec.* 1986, pp. 364 - 374
2. **R. Klette, D. Mehren, V. Rodehorst**, An Application of Shape Reconstruction from Rotational Motion, *Real-Time Imaging* (1995), to appear
3. **A. Koschan**, A Framework for Area-Based and Feature-Based Stereo Vision. *Machine Graphics & Vision* 2 (1993), pp. 285 - 308
4. **Y. Shirai, Y. Nishimoto**, A Stereo Method Using Disparity Histograms of Multi-Resolution Channels, *Proc. 3rd Int. Symp. on Robotics Research*, Gouvieux, France, 1985, pp. 27 - 32
5. **A. Koschan**, Stereo Matching Using a New Local Disparity Limit, *Proc. IVth Int. Conf. Computer Analysis of Images and Pattern*, Dresden, September 17 - 19, 1991, pp. 48 - 53
6. **A. Koschan**, How to Utilize Color Information in Dense Stereo Matching and in Edge-Based Stereo Matching. *Proc. 3rd Int. Conf. Automation, Robotics and Computer Vision*, Singapore, November 8 - 11, 1994, Vol. 1, pp. 419 - 423
7. **Y.-I. Ohta, T. Kanade, T. Sakai**, Color Information for Region Segmentation. *CGIP* 13 (1980), pp. 222 - 241
8. **J.L. Barron, D.J. Fleet, S.S. Beauchemin**, Performance of Optical Flow Techniques, *Int. J. Computer Vision* 12 (1994), pp. 43 - 77
9. **R. Klette, P. Handschack**, Quantitative Comparisons of Differential Methods for Measuring of Image Velocity. in: C. Arcelli, L.P. Cordella, G. Sanniti Di Baja (eds.), *Aspects Of Visual Form Processing.* World Scientific, Singapore 1994, pp. 241 - 250
10. **R.J. Woodham**, Photometric Method for Determining Surface Orientations from Multiple Images, *Optical Engineering* 19 (1980), pp. 139 - 144
11. **R.T. Frankot, R. Chellappa**, A Method for Enforcing Integrability in Shape from Shading Algorithms. *IEEE Trans. on PAMI* 10 (1988), pp. 439 - 451
12. **G. Bellaire, K. Schlüns, A. Mitritz, K. Gwinner**, Adaptive Matching Using Object Models Generated from Photometric Stereo Images. *Proc. 8th Int. Conf. on Image Analysis and Processing*, San Remo, Italy, Sept. 13 - 15, 1995, to appear
13. **B.K.P. Horn**, *Robot Vision*, Mcgraw-Hill, New York 1986
14. **G.J. Klinker, S.A. Shafer, T. Kanade**, A Physical Approach to Color Image Understanding. *IJCV* 4 (1990), pp. 7 - 38
15. **K. Schlüns, O. Wittig**, Photometric Stereo for Non-Lambertian Surfaces Using Color Information. *Proc. 7th Int. Conference On Image Analysis And Processing*, Monopoli, Italy, Sept. 20 - 22, 1993, pp. 505 - 512.
16. **A.M. McIvor**, Three-Dimensional Vision Applied to Natural Surfaces. Industrial Research Limited Report 325, Auckland, October 31, 1994
17. **R. Klette, A. Koschan, K. Schlüns, V. Rodehorst**, Surface Reconstruction based on Visual Information, TR95/6, July 1995, Dep. of Computer Science, The University of Western Australia