



HAL
open science

Voxel Carving for Specular Surfaces

Thomas Bonfort, Peter Sturm

► **To cite this version:**

Thomas Bonfort, Peter Sturm. Voxel Carving for Specular Surfaces. 9th IEEE International Conference on Computer Vision (ICCV '03), Oct 2003, Nice, France. pp.691-696, 10.1109/ICCV.2003.1238401 . inria-00525642

HAL Id: inria-00525642

<https://hal.inria.fr/inria-00525642>

Submitted on 26 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Voxel Carving for Specular Surfaces

Thomas Bonfort and Peter Sturm

MOVI - GRAVIR - INRIA, Grenoble, France

{thomas.bonfort, peter.sturm}@inrialpes.fr

Abstract

We present an novel algorithm that reconstructs voxels of a general 3D specular surface from multiple images of a calibrated camera. A calibrated scene (i.e. points whose 3D coordinates are known) is reflected by the unknown specular surface onto the image plane of the camera. For every viewpoint, surface normals are associated to the voxels traversed by each projection ray formed by the reflection of a scene point. A decision process then discards voxels whose associated surface normals are not consistent with one another. The output of the algorithm is a collection of voxels and surface normals in 3D space, whose quality and size depend on user-set thresholds. The method has been tested on synthetic and real images. Visual and quantified experimental results are presented.

1. Introduction

3D shape reconstruction techniques obtain models of real-world objects, that can then be used in computer graphics, CAD, multimedia databases, etc... Most reconstruction methods rely on the identification and matching of pixels corresponding to a same object feature, and return the 3D coordinates of the object's feature using different geometric constraints. In the case of specular objects, these non-specific methods will always fail to reconstruct the object's surface. Indeed, the observed texture of a specular object is the reflection of the object's surrounding environment, rather than the texture of the object itself. As the viewpoint changes, the observed texture *moves* along the object's surface, thus invalidating the geometric constraints used by classical reconstruction methods.

We describe a method recovering points of a specular surface. Texture and shading contributions of the surface are ignored, our focus being only on perfect mirrors like objects made of polished metal. We place ourselves in the case of several views of calibrated cameras observing the specular surface, the images seen by these cameras being the reflection of the object's surrounding environment. We assume that this environment (typically a printed target, from now on referred to as *scene points*) is calibrated, i.e. we know the 3D coordinates of a number of scene points.

A realworld application of our reconstruction method

uses a printed target attached to a camera taking images of a specular surface. Camera pose is obtained by taking images of the system from a stereo rig. Once the matching of the points on the target with the image of their reflection has been accomplished, a set of voxels belonging to the surface is obtained by our method, along with their corresponding surface normals. A model of the object could then be calculated by fitting a surface to the obtained points and normals.

1.1. Previous Work

Specularities have interested researchers in the field of computer vision for the past 20 years. For example, Blake *et al.* [1] studied the disparity of highlights on a specular surface seen from two viewpoints. Zisserman *et al.* [17] tracked the motion of specularities to obtain information on the surface. Healey [5] used static images and a reflectance map to recover 3D points on a specular surface. In [8], Oren and Nayar studied the classification of real and reflected features, and recover the profile of a specular surface by tracking an unknown scene point. The profile can be recovered without further hypotheses only if the motion of the camera, the scene point and its reflection in the mirror surface are coplanar, thus limiting practical applications. Halstead *et al.*, in [3], fit a spline surface to a set of normals, iteratively refining the result. Their method requires an initial seed point on the specular surface, and was applied to the sub-micronic reconstruction of the human cornea. Schultz recovers in [11] the ocean's surface given three calibrated images, an irradiance map of the illumination, and known seed points. An elevation map is obtained by propagating around the known points using observed surface normals, while minimising the difference between real and rendered images. In [9], Ripsman and Jenkin recover specular planes using three cameras and active illumination. The aimed application is the automatic inspection of orbital objects. Savarese and Perona detail in [10] the information available for one view of a specular object reflecting three or more intersecting calibrated lines. The second order surface geometry can be obtained up to one unknown parameter. In the case of general specular surfaces, Zheng and Murata in [16] reconstruct a rotating specular object by studying the motion of the illumination created by two circular light sources.

The method we propose makes no assumptions on the

continuity or regularity of the specular surface, and doesn't need any initial seed points to complete. It extracts voxels of the surface independently from one another, therefore preventing the accumulation of errors that can occur in other methods.

2. Geometric Constraints

This paragraph presents the basic geometric constraints used by our method. The following notation will be used throughout the article: **bold** letters represent a vector in 3D space, while *italic* letters represent scalars. A $\hat{\cdot}$ represents a normalized vector. Subscripts are used as follows: O_c is the *camera's* projection centre, x_s is a calibrated *scene* point, x_m is the position of the reflection of x_s on the *mirror* surface (therefore it depends on the viewpoint), x_i is the projection of x_m on the *image* plane of the camera.

- **Ideal Specular Surfaces :** We have chosen to consider only purely specular surfaces such as objects made of polished metal. With this assumption, the law of reflection links the surface normal \hat{n}_m , the incoming light direction \hat{r}_i , and the reflected light ray \hat{r}_r by $\hat{r}_r = \hat{r}_i - 2(\hat{n}_m \cdot \hat{r}_i)\hat{n}_m$. In other words, the surface normal is the bisector of the angle formed by the incoming and outgoing light rays.
- **Image Formation :** The formation of an image of a specular surface is shown in figure 1: given a known scene point x_s , a known mirror point x_m , and a known image point x_i , the surface normal n_m at x_m is constrained by:

- n_m belongs to the plane formed by x_s , x_m and the projection ray formed by O_c and x_i .
- the angle α_i between the incident line of sight and the surface normal is equal to the angle α_r between the surface normal and the reflected line of sight.

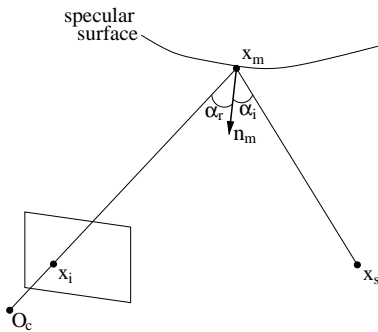


Figure 1: **Image formation.** A calibrated scene point is reflected by the specular surface onto the image plane of a calibrated camera.

Thus, the surface normal at x_m is given by the bisector of the angle formed by the scene point, the voxel and the camera. This constraint is on its own clearly insufficient to determine the surface's position and orientation: given

any point on the projection ray formed by the camera's projection centre O_c and the image point x_i , we can find the orientation of a specular surface passing by that point that would lead to the same observation. Inversely, given a surface normal, the position of a surface leading to the same observation can be obtained. We therefore need further constraints so as to find surface position and orientation simultaneously, which we obtain by using multiple images of the specular surface and the fact that the normal at a given point on a specular surface is independent of the viewpoint.

3. Reconstruction Method

This section presents our reconstruction method, using multiple views of a calibrated camera. We start by discretizing the 3D space around the specular object, to obtain a voxel representation of the working space. The reconstruction then takes place in two main phases: the first phase associates $n \geq 0$ normals to every voxel of the 3D space surrounding the specular surface. In the second phase, along each projection ray, the voxel whose associated normals are the most consistent with one another is kept. In essence, our method is very similar to voxel carving techniques [12, 7] who rely on surface color to reject incorrect voxels.

3.1. Phase 1: Normal Accumulation Process

Figure 2 shows the information available for one viewpoint, once the 3D space has been discretized: for each voxel traversed by the projection ray corresponding to an image point and a scene point, we compute the surface normal associated to that voxel. This surface normal corresponds to the normal of a specular plane passing through the centre of the voxel, that would produce the same observation. In other words, for a given scene point x_s reflected by the specular surface onto the image plane at x_i , we associate for every voxel x_m traversed by the projection ray $[O_c x_i)$ the normal of a specular plane passing by the centre of that voxel, that would reflect x_s onto x_i .

The normal calculation is repeated for every voxel traversed by every projection ray corresponding to the reflection of the scene points, and for every camera position. Each voxel in the 3D space surrounding the specular object will then have $n \geq 0$ surface normals associated to it.

3.2. Phase 2: Discarding Incorrect Voxels

Let us consider the voxels traversed by the projection ray formed by the reflection of the scene point x_s by the specular point x_m onto the image point x_i . There are two cases occurring for a computed normal at a given voxel:

- **the voxel does not belong to the specular surface:** In this case, the associated normal has no physical reality. If the voxel is traversed by a second ray originating from another viewpoint, the second associated normal has no reason to be similar to the first one.

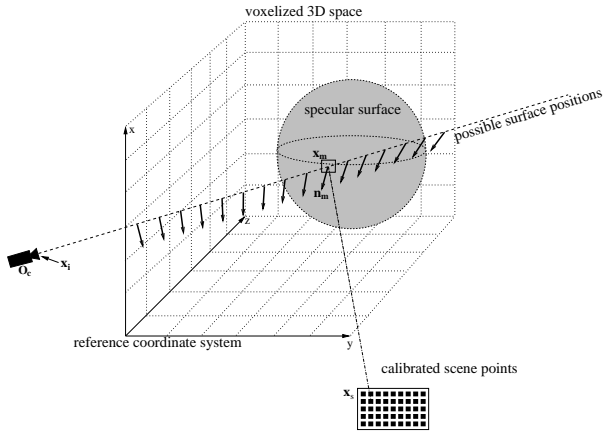


Figure 2: Normal construction for each voxel traversed by a projection ray. The process is repeated for each calibrated scene point x_s , and for each camera position O_c .

- **the voxel belongs to the specular surface:** In this case, assuming the voxel discretization is infinitely precise and there is no noise in the measures of x_s and x_1 , the associated normal is the normal of the specular surface at that point. If the voxel is traversed by another projection ray, the voxel's second associated normal will be identical to the first one. Intuitively, voxels belonging to the surface can be selected by thresholding the disparity of their associated normals, as illustrated in figure 3. Furthermore, as for each viewpoint a single correct voxel corresponds to each projection ray, incorrect voxels can be discarded by stepping along each projection ray and keeping the voxel with the least disparity.

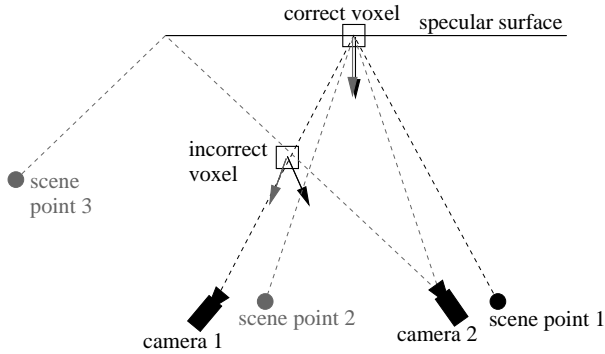


Figure 3: Normals associated to a correct and incorrect voxel. The normals associated to a voxel that does not belong to the specular surface are dissimilar, whereas those associated to a voxel belonging to the surface are consistent with one another.

• **Quantifying Normal Disparity:** So as to discard incorrect voxels, a quantifiable disparity measure is necessary. The simplest disparity measure is given by the mean angle

between each normal and the mean normal:

$$disparity = \frac{\sum_{i=1}^n \arccos^2(\hat{n}_i \cdot \bar{\hat{n}})}{n} \quad (1)$$

where $\bar{\hat{n}}$ stands for the mean normal associated to a voxel, *i.e.*

$$\bar{\hat{n}} = \frac{\sum_{i=1}^n \hat{n}_i}{\left\| \frac{\sum_{i=1}^n \hat{n}_i}{n} \right\|}$$

This disparity measure is in practice insufficiently discriminant. As well as normal disparity, depth information must be taken into account: as illustrated in figure 4, the distance of a voxel to the cameras plays an important role in the disparity associated to a voxel's normals. Two different approaches have been used to obtain correct results, as explained in the following paragraphs.

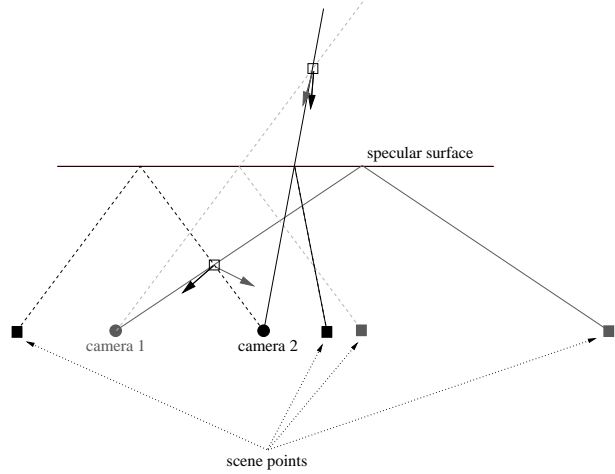


Figure 4: Disparity variation due to camera distance. Two equally incorrect voxels have an associated normal disparity that depends on their distance to the cameras. A threshold-based decision will tend to keep incorrect voxels if they are further away from the camera viewpoints.

• **Reprojection Error :** Additional information is available if we compare the reconstructed surface to the images we have of it, by checking that the images of every scene point reflected by a given voxel (whose surface normal is the mean normal seen in the previous paragraph) are close in the participating cameras to the images of the same scene point reflected by the original surface.

A given voxel is assumed to be a plane passing by the centre of the voxel, whose normal is the voxel's mean normal. Every participating scene point associated to the voxel is then geometrically reflected by this plane onto the image plane of the participating cameras. The reprojection error measure becomes the mean square distance of the reflections to the original pixel.

This error measure does not take normal disparity into account, and therefore is not sensitive to depth variations.

Qualitatively correct reconstructions have been obtained with this method, but even better results have been obtained with the following heuristics.

- **Heuristic Disparity Measures :** We have tested several heuristic disparity measures that take depth information of the voxels into account, and the best results have been obtained by dividing the disparity obtained in (1) by the mean angle formed by the scene points, the voxel, and the different camera viewpoints. As these angles tend to decrease as the voxels are further away from the camera viewpoints, this disparity measure penalises the furthest voxels.
- **Limitation :** The reconstruction of a specular object is not straightforward: figure 5 shows that a projection ray that intersects the object’s surface twice will cause correct voxels to be associated with incorrect normals, thus invalidating a decision process based on normal disparity. This imposes that the reconstruction take place in several steps, or that a robust disparity measure be used: a correct voxel will be associated with a number of inconsistent normals originating from a double intersection, plus a number of identical normals corresponding to the surface normal. Thus, this limitation could be overcome by selecting consistent normals using robust statistics, and applying the disparity measure on these selected normals. The results we present in the next sections were obtained by placing the cameras in positions assuring no double intersections took place, therefore reconstructing a patch on the specular surface.

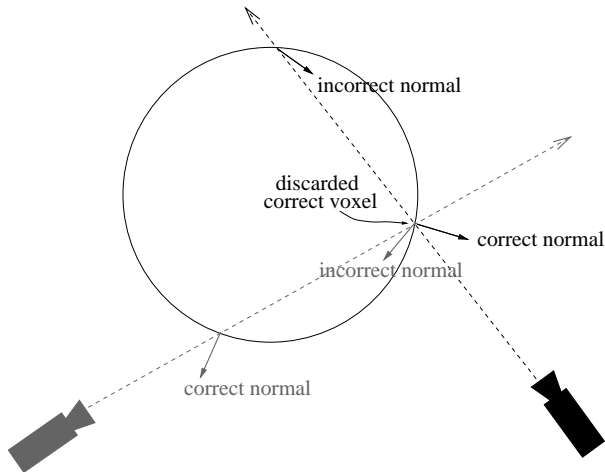


Figure 5: double intersections with the object’s surface cause correct voxels to be associated with inconsistent normals, and therefore be discarded when measuring their disparity.

4. Results on Simulation Data

We have tested our method on the reconstruction of specular spheres and planes, using simulation data and ray-traced images. We deliberately limited ourselves to these simple

objects so as to easily and automatically perform the correspondence problem between the scene points and the image points.

- **Setup :** The results shown in the next paragraph were obtained using 40 ray-traced images of a specular sphere. A black on white checkerboard target was reflected by the sphere onto the image plane of the cameras, and a Harris corner detector [4] was used to extract the corners in the obtained images, thus creating some noise in the measures. The extracted points were then analytically matched with their corresponding scene points on the target, using the known geometry of the sphere.
- **Quantified Results :** We have tested our reconstruction method with different thresholds, as shown in table 1, using the heuristic disparity measure seen above. The quality of the output depends on the minimum number of normals a voxel must be associated with before the disparity measure can be trusted. With less than 4 normals per voxel, a number of outliers can appear. With at least 8 normals, the extracted voxels are never further than two voxel to the original surface.

	strict	medium	laxist
4	363 93.4%	969 95.3%	3103 94.45%
8	161 100%	589 100%	1982 100%

Table 1: Results obtained using strict, medium and laxist thresholds on normal disparity. The lines represent the minimum number of normals associated to a voxel before this voxel can be considered. For each combination, the first number is the number of accepted voxels, and the second one is the percentage of these voxels that are either intersecting the specular surface or less than two voxels away from an intersecting voxel.

5. Results on Real Images

We have tested our method on the reconstruction of a part of a specular spoon made of low end polished metal, thus exhibiting consequent normal variation on some parts (see figure 6, a little left from the center for example). In the next paragraphs we present the experimental setup, detail the calibration process which is non-trivial due to the limited information available, and present a couple of visual results of the reconstruction.

5.1. Experimental Setup

Figure 7 shows the experimental setup. A color coded [6] printed target was attached to an Olympus C2500 digital camera taking images of the specular spoon from 56 different viewpoints, while a stereo rig was used to obtain the pose of the target + camera system for every viewpoint. The color code uniquely locates a circle on the target given its

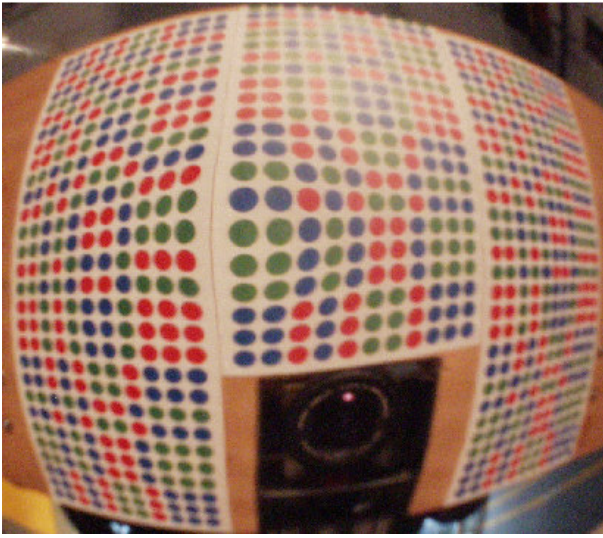


Figure 6: One of the 56 images used for reconstruction. Around 2 out of 3 colored circles were automatically extracted for this image. The color code uniquely identifies a circle given its color and the color of its 4-neighbours.

color and the colors of its 4-neighbours, thus allowing us to automatically match pixels in the image to the corresponding points on the target. We then extracted 1890 voxels corresponding to the specular surface out of a $100 \times 100 \times 100$ cube of voxels, each of these voxels being 1 mm^3 . We wish to point out that we used low end imaging devices, and that the automatic segmentation of the images for feature extraction induces consequent noise in the location of the pixels corresponding to the reflection of the target points. Therefore, pose estimation of the printed target and the camera at the different viewpoints, and projection lines corresponding to pixels were unlikely to be very accurate, making us believe that this reconstruction method is relatively robust to medium noise.

One point that should be mentioned is blur. The images we used exhibited large out of focus regions, thus complicating feature extraction on the specular surface. While this defocus did not prevent the reconstruction, having multiple images with a different focus setting, for every viewpoint, would be a simple method to resolve the problem.

5.2. Calibration

The main device used is the combination of the digital camera and the target plane attached to it, whose reflection in the specular object is what the camera sees. For each image acquired using this device, we need to know the (relative) pose of both, camera and target plane. Since they are rigidly linked, the problem reduces to determining the pose of either one of them for the current image, and the relative pose between them, that is fixed.

Several potential solutions come readily to mind. For

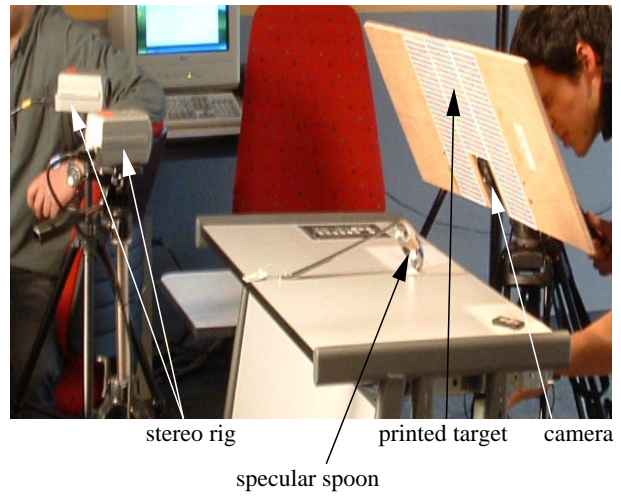


Figure 7: Experimental setup. The printed target is reflected by the specular sphere onto the image plane of the camera. The system's pose is determined by the stereo rig.

example, we might put the specular object on a planar platform with targets printed on it. The camera's pose could then be computed using these targets. We rejected this solution because due to the specular object's convex form, the camera had to zoom deeply onto the object, in order to extract the reflected patterns in the image. So, it was not sure that targets on the platform would be systematically visible.

We thus chose to use a fixed stereo system, placed behind the specular object and that observes the camera + target device (see figure 7). The stereo system gives us the pose of the target plane for each acquisition position (the method of [14] was used). Actually, the stereo system was calibrated (intrinsic and relative pose) using directly the images of the target plane acquired during the experiments (using the methods described in [15, 13, 14]).

The remaining problem, estimating the relative position of the camera and the target plane, is not trivial, since the camera has no direct view of the targets. We solved the problem in a way analogous to [2]. Instead of the specular object, we let the camera observe a 3D calibration grid. A dozen images were acquired. The camera's and the target plane's ego-motions across the viewing positions, were computed as follows. For the camera, this is done using the 3D calibration grid and a classical camera calibration + pose algorithm. As for the target plane, its pose and thus ego-motion is output by the stereo system, as described above. The two sequences of ego-motions were then aligned using a method inspired from [2] and whose formulation is closely related to classical hand-eye calibration. This alignment transformation allows finally to compute the rigid transformation that links the target plane and the camera.

Developing this calibration process was rewarding in itself, but details have to be omitted due to lack of space.

5.3. Results

Reconstruction results for a patch of the specular spoon are shown in figures 8 and 9. What these images do not clearly show but can be noticed in the 3D model is that there are no holes greater than 1-2 voxels on the patch, and that the extracted thickness is around 1-3 voxels. The extracted voxels are qualitatively correct, and quantitatively correct up to what we could measure.

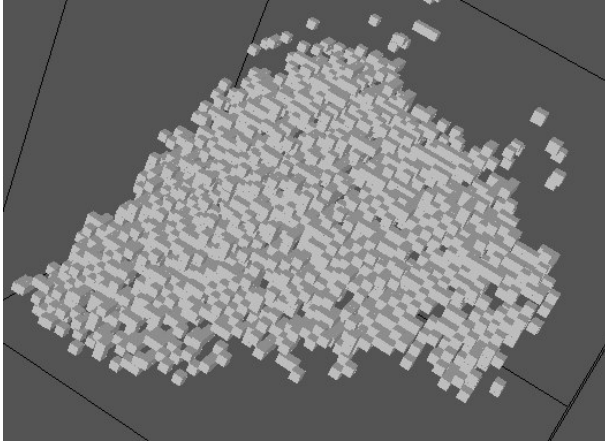


Figure 8: Front view of the reconstructed voxels.

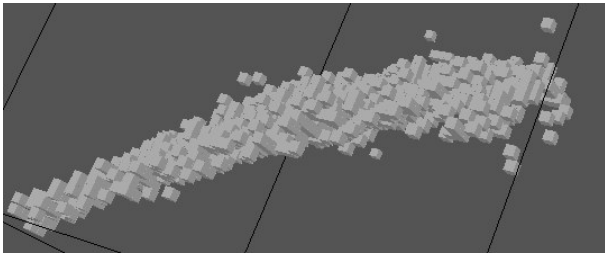


Figure 9: Side view of the reconstructed voxels. The spoon's curvature is clearly visible in one dimension. Views from other directions show curvature is correct on all the reconstructed surface.

6. Conclusion and Future Work

We have developed a method reconstructing voxels and normals of a specular surface, requiring multiple views of a calibrated camera and the matching of calibrated scene points with their reflection in the specular surface. We have tested this method on simulated data and obtained qualitatively correct results, and on real images and obtained correct results as far as we could measure.

The underlying foundation is based on the consistency of surface normals whatever the viewpoint. A reprojection error measure and several heuristic disparity measures have been tested to extract surface voxels, one of the latter giving the best results. The principal advantage we see of the method is that it makes no assumption on the surface, and

does not need any initial seed points to complete. It is therefore not subject to error accumulation that can occur when assumptions on surface continuity or derivability are made.

An interesting challenge still remains the information available for specular surfaces from uncalibrated views and/or an uncalibrated reflected scene.

References

- [1] A. Blake and G. Brelstaff, Geometry from Specularities, *ICCV*, p 394–403, 1988
- [2] Y. Caspi and M. Irani, Alignment of Non-Overlapping Sequences, *ICCV*, pp. 76-83, 2001
- [3] M. Halstead, B. Barsky, S. Klein and R. Mandell, Reconstructing Curved Surfaces from Specular Reflection Patterns using Spline Surface Fitting of Normals, *SIGGRAPH*, 1996
- [4] C.G. Harris and M. Stephens, A Combined Corner and Edge Detector, *4th Alvey Vision Conference*, p 147–151, 1988
- [5] G. Healey and T. Binford, Local Shape from Specularities, *Computer Vision, Graphics, and Image Processing* **42**, p 62–86, 1988
- [6] P.M.Griffin, S. Narasimhan and S.R. Yee, Generation of Uniquely Encoded Light Patterns for Range Data Acquisition, *Pattern Recognition* **6**,p 609–616, 1992
- [7] K. N. Kutulakos and S. M. Seitz, A Theory of Shape by Space Carving, *International Journal of Computer Vision*, 38(3), pp 199-218, 2000
- [8] M. Oren and S. Nayar, A Theory of Specular Surface Geometry, *International Conference on Computer Vision*, p 740- , 1995
- [9] A. Ripsman and M. Jenkin, Local Surface Reconstruction of Objects in Space, *IEEE Int. Symposium on Computational Intelligence*, 2001
- [10] S. Savarese and P. Perona, Local Analysis for 3D Reconstruction of Specular Surfaces – Part II, *European Conf. on Computer Vision*, 2002
- [11] H. Schultz, Shape Reconstruction from Multiple Images of the Ocean Surface, *Photogrammetric Engineering and Remote Sensing*, vol. 62, no. 1, 1996
- [12] S. M. Seitz and C. R. Dyer, Photorealistic Scene Reconstruction by Voxel Coloring *International Journal of Computer Vision*, 35(2), pp 151-173, 1999
- [13] P. Sturm and S. Maybank, On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications, *CVPR*, pp. 432-437, 1999
- [14] P. Sturm, Algorithms for Plane-Based Pose Estimation, *CVPR*, p.. 1010-1017, 2000
- [15] Z. Zhang, A Flexible New Technique for Camera Calibration, *IEEE PAMI*, Vol. 22, No. 11, pp. 1330-1334, 2000
- [16] J. Zheng and A. Murata, Acquiring a Complete 3D Model from Specular Motion under the Illumination of Circular-Shaped Light Sources, *IEEE Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, 2000
- [17] A. Zisserman, P.Giblin, and A. Blake, The Information Available to a Moving Observer from Specularities, *Image and Video Computing* **7**, p 38–42, 1989