Global Depth from Epipolar Volumes - A General Framework for Reconstructing Non-Lambertian Surfaces

Timo Stich[1], Art Tevs[2], Marcus Magnor[1]

Computer Graphics Lab[1]
TU Braunschweig, Germany
stich,magnor@cg.cs.tu-bs.de

MPI Informatik[2]
Saarbrücken, Germany
tevs@mpi-inf.mpg.de

Abstract

Using Epipolar Image Analysis in the context of the correspondence finding problem in depth reconstruction has several advantages. One is the elegant incorporation of prior knowledge about the scene or the surface reflection properties into the reconstruction process. The proposed framework in conjunction with graph cut optimization is able to reconstruct also highly specular surfaces. The use of prior knowledge and multiple images opens new ways to reconstruct surfaces and scenes impossible or error prone with previous methods. Another advantage is improved occlusion handling. Pixels that are partly occluded contribute to the reconstruction results. The proposed shifting of some of the computation to graphics hardware (GPU) results in a significant speed improvement compared to pure CPU-based implementations.

1 Introduction

Solving the passive multi-view 3D reconstruction problem has and still is one of the most worked on problems in the computer vision community. Motivated by the human ability to easily perceive a 3D world with only two "cameras", many different approaches have been developed. However, most approaches rely on strong assumptions on the BRDF [1] of the scene objects e.g. to be lambertian which is not true in general. For example shiny materials like plastic or metals violating this assumption cause artifacts and errors in the reconstruction results. In our work we formulate the correspondence finding problem in terms of the Epipolar Image analysis (EPI). The use of prior knowledge about reflectance properties in general opens up ways to reconstruct surfaces and scenes impossible or error prone with previous methods. The main contribution of the presented approach is the formulation of the dense depth map estimation problem when prior knowledge about reflectance properties of the surfaces is available.

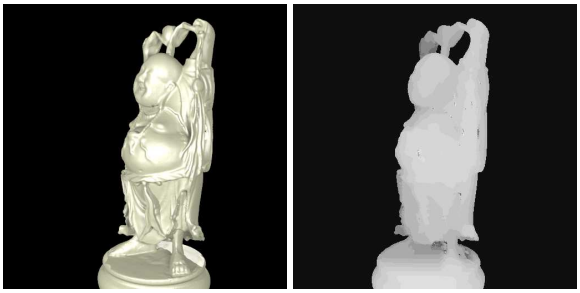

Figure 1: The left image shows the rendering of the Happy Buddha model [4] with a plastic material as seen from camera 5. 10 different views are used in total for reconstruction. The right image shows the reconstruction results obtained with the proposed method. The statue is illuminated with three area light sources that cause multiple specularities and interreflections. However, the reconstruction result contains almost no artifacts.

We pose the reconstruction process as an energy minimization problem and solve it via the well known graph cut algorithm, first used in this context by [2] and [3]. For the implementation we shifted some computations to the Graphics Hardware (GPU) achieving significantly improved runtimes over a CPU only implementation. Some results and a comparison to the results of previous methods are presented (see e.g. Figure 1). Since the proposed method is a passive multi view reconstruction approach the depth map estimation can be used to analyze dynamic scenes as well.

2 Related Work

There are numerous publications in the field of multi-view 3D reconstruction. Most of the work done so far, assumes that the BRDF of the surface is lambertian meaning specular or anisotropic reflectance behavior causes errors in the reconstruction result.

Like some of the works presented before, we also use graph cut optimization to find a dense depth map estimation. Applying graph cuts to solve the correspondence finding problem for multiple views handling occlusion was first proposed in [3]. The data term used in this work is however based on the color constancy principle, which becomes problematic for non-lambertian surfaces. Occlusion is also only considered between neighboring images and is implemented as a check that prevents estimating the depth of partly occluded pixels. In our approach we implemented occlusion handling directly in the data energy term. Being less conservative, our proposed occlusion handling finds results for partially occluded pixels. Another related work using graph cuts was presented in [5]. It is an extension of the multi-view graph cut algorithm to include the detection of background into the reconstruction process. However, handling of specular surfaces is not covered.

Another work applying the graph cut algorithm in this field was presented by Davis et al. [6]. They achieve good depth estimations of nearly arbitrary BRDF surfaces using controlled variation of the scene illumination assuming light transport constancy. However, the depth map estimation only works for static scenes, since multiple images with different lighting are needed. Another approach to handle specularities is to remove specular pixels from the reconstruction process [7, 8] or to handle outliers and occlusions via hidden variables in the reconstruction process [9].

Other approaches for multi-view 3D reconstruction work by finding the scene geometry rather than computing depth maps. A prominent example is the Space Carving approach [11]. Here the space is discretized into voxels which are then sequentially tested for color constancy. There are also numerous extensions to this approach e.g. [12, 13]. The first work uses priors to improve the results and the second deals with specular highlights and textureless regions. Lately, a new approach using Surfels to both reconstruct the surface of an object and view independent reflectance maps [14] was introduced. Surfels are also used in [10]. The authors propose a voting approach to estimate the parameters of the Phong BRDF model to account for the reflection properties of the surface. The discretization of the resulting geometry however introduces additional artifacts into the model based results. A different approach for computing the underlying model for lambertian plus specular surfaces is the method published in [15]. Here the reconstruction for non-lambertian surfaces is achieved via a rank constraint on the radiance tensor.

Our algorithm uses EPI analysis, specifically Epipolar Volumes built from a multi camera setup to perform depth estimation. This has first been introduced in the context of structure from motion by [16]. Recent work by [17] expanded the approach to handle occlusions and specularities. However, their approach does not estimate the depth of specular pixels but tries to handle them as invalid outliers.

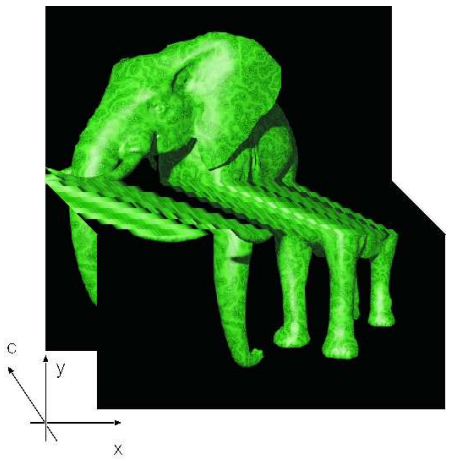

Figure 2: Cut through the Epipolar Volume V created from the set of rectified images $\mathbf{I}$. The images are stacked so that the xy-plane at $c = i$ is defined by image I_i, with $i \in \{1 \dots N\}$.

3 Problem Statement

Given an *Epipolar Volume* constructed from rectified camera images of a scene (e.g. Figure 2), we reconstruct depth maps incorporating any knowledge available about the reflectance properties of objects found in the scene and assumed piecewise smooth surfaces. Rectified meaning that scanlines of the images correspond to epipolar lines of the multi camera recording setup . This problem can be posed as a minimization of an energy functional on lines defined in the Epipolar Volume and a neighborhood regularization. In a second step we show how to add occlusion handling as a straight forward extension of the proposed algorithm.

Assume N rectified images $\mathbf{I}$ of a scene taken simultaneously from N different viewpoints on the same baseline. Intuitively speaking, from the image set $\mathbf{I}$ we can then create an epipolar volume V as shown in Figure 2 by defining a new 3D space using the image coordinates x, y and the camera position c on the baseline as the third dimension. Given a point P in the scene, this point is then found on a line l_P in V. The important observation is that the slope λ_p of l_P is inversely proportional to the depth of P. Thus if $\lambda = 0$, $P_z = \infty$. Note that λ is equivalent to *disparity* in the special case of $N = 2$.

The color values on l_P are defined by the BRDF of P, $f_r(\theta_i, \phi_i, \theta_r, \phi_r)$. Thus we get

$$l_P(i) = f_r(\theta_i(i), \phi_i(i), \theta_r(i), \phi_r(i)) \; \forall i \in \{1 \dots N\} \quad (1)$$

For the line l_P describing P in the Epipolar Volume the following assumptions hold. The lighting setup for the scene,

the local coordinate system for each P and the associated BRDF are constant. Thus the variations in the color values for the observations of P are dependant on the viewing directions only. Applying assumptions or prior knowledge about this variations to build a vector $\hat{f}_r$ then allows to pose the problem of estimating the depth of P as minimizing an energy functional of the form

$$E(p, \lambda) = d(l_P, \hat{f}_r) \tag{2}$$

where $d(\cdot, \cdot)$ is a distance measure between two vectors. Finding λ_P^* is achieved by solving

$$\lambda_P^* = \arg\min_{\lambda} E(p, \lambda) \tag{3}$$

The dense depth map for a given image is then built by solving Equation 3 for each pixel in the image under some neighborhood regularization constraints. In particular we use the assumption that the BRDF varies smoothly at λ_P^* to reconstruct also non-lambertian surfaces.

4 Epipolar Volumes

For a scene at time t, the radiance observed at point x from a given direction r is described by the Plenoptic function [18] $O(x, r)$; $O : \mathbb{R}^3 \times \mathbb{S}^2 \rightarrow \mathbb{R}^d$ where $\mathbb{S}^2$ is the unit sphere of directions in $\mathbb{R}^3$ and $d = 1$ for intensity or $d = 3$ for color images. Assuming the pinhole camera model, an image $I(c) = O(c, r)$ is thus defined by the Plenoptic function with the camera center c over a subset of $\mathbb{S}^2$ which is determined by the field of view and the viewing direction v of the camera.

Given camera centers $c \in l_c$ with

$$l_c = c_0 + \alpha d \tag{4}$$

with $d \perp v$ and without loss of generality $\alpha \in \mathbb{R}^+$. We define the Epipolar Volume $V \in \mathbb{R}^2 \times \mathbb{R}^+$ as

$$V = I(l_c) = O(l_c, r) \tag{5}$$

as a subset of P.

Using this definition, images of a scene created from cameras placed on the same baseline l_c are samples of V at discrete points on the c-axis. Specifically, we define l_c as the positive c-axis of the Epipolar Volume resulting in a sampled Volume as shown in Figure 2.

As described, all cameras are placed on l_c and have the same viewing direction v. The projections of a scene point P vary only in one dimension, which is parallel to l_c since the images are then rectified by definition [19].

4.1 Depth

Assume a point $P \in \mathbb{R}^3$ in the scene visible from all cameras c_i. Let the points $p_i \in \mathbb{R}^2$ be the projections of P in $I(c_i)$. As introduced before every p_i corresponds to a direction from c_i to P. Then the lines defined by c_i and the direction associated with p_i intersect in P (see Figure 3).

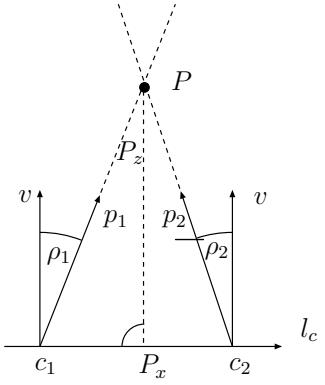

Figure 3: Triangulation of a point P in the scene. The rays defined by p_1 and p_2 intersect at P. The distance between P and the camera plane defined by v and l_c is denoted as P_z

To explore the distribution of the p in V we consider the plane that is formed by l_c and P. In this plane, the direction p reduces to the angle ρ and P_x, $P_z > 0$ describe the position of P. We find

$$\tan \rho = \frac{P_x - c}{P_z} \tag{6}$$

for any point c on l_c. Given two camera centers c_1 and c_2 on l_c with $c_1 = c_2 + \delta,\ \delta > 0$ we define

$$\lambda = \tan \rho_2 - \tan \rho_1 = \frac{\delta}{P_z} \tag{7}$$

With definition (5) and setting $\delta = 1$ we conclude that c_i is the c-axis coordinate in V and thus the p form a line l_P in V where the slope λ of l_P is inversely proportional to the distance P_z between P and the camera base line.

$$l_P = \lambda(c + P_x) \tag{8}$$

Conversely, the depth of a point P is defined by l_P in V. Searching for l_P in V and computing λ determines the depth of P. In our implementation we exploit this property to pose the correspondence search problem as an energy minimization. Since we use graph cuts to solve the NP-hard problem, we have to discretize the possible depth solutions first. This is done by specifying an upper and lower limit for λ and interpolating in between.

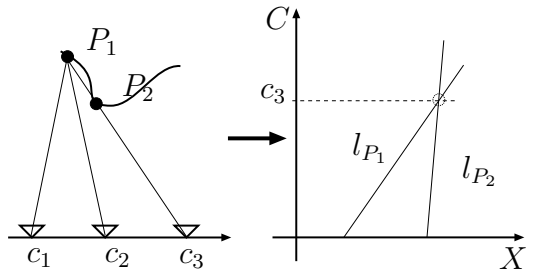

Figure 4: Occlusion in the Epipolar Volume. Since the slope λ is inversely proportional to the depth of a surface point P, P_2 occludes P_1 at the intersecting camera c_3. Thus when extracting l_{p1} the according value is invalidated.

4.2 Occlusion

Pixels that are occluded in some views violate the assumption that P is found on line l_P in the Epipolar Volume. However, occlusions in Epipolar Volumes have well defined properties. This allows to estimate depth values also for partially occluded pixels. If a point P_1 is occluded by a point P_2 then the following assumptions hold (see Figure 4).

1. P_2 has a smaller distance to the camera baseline than P_1
2. The corresponding lines l_{P_2} and l_{P_1} intersect at the occlusion

From (7) we conclude the criterion $\lambda_1 < \lambda_2$ can be used for occlusion checking. Let the set of points p_{occluded} define the points on l_P that are occluded as defined above. Then we define

$$\hat{l}_P = l_P \setminus p_{\text{occluded}} \tag{9}$$

as the line with occluded pixels removed. The assumption that P is defined by $\hat{l}_P$ then also holds in the case of occlusion.

In the reconstruction process we implement occlusion detection in the data term of the energy function. From the current solution, we can check the depth neighborhoods for each pixel in the Epipolar Volume and decide if the pixel is occluded in one or more images. If occluded, the pixel value is invalidated for the occluded images and accordingly handled in the energy computation. Partly occluded pixels are thus considered in depth reconstruction while depth values of pixels visible in one view only are extrapolated from their neighbors.

4.3 BRDF Models

The main contribution of this work is the way prior knowledge and assumptions about the BRDFs found in the scene are introduced in the reconstruction process. Utilizing prior knowledge significantly improves the quality of the reconstruction result and allows reconstructing surfaces with a wider range of materials than previously possible.

The Epipolar Volume defines a function l_P, describing the color value variation dependend on the camera position c. In this section we focus on how these values vary along the l_P. Generally the values of a point P are defined by its BRDF. The BRDF is itself dependent on the local coordinate system associate with P. For the Epipolar Volume at a given point in time, we find that the local coordinate systems, the BRDF and the lighting setup L are constant for all l_P. Thus the values depend only on the different viewing directions p_i. This leads to the conclusion that the values on l_P are samples of the BRDF over the viewpoints under a given lighting setup.

In this work we explore the potential of our proposed approach to the calculation of dense depth maps for highly specular and anisotropically reflecting surfaces. We propose the assumption that the BRDF of any point P varies smoothly in the range of l_c. This already gives good results for shiny objects made of metals and plastic. There is further no connection on BRDFs of neighboring points which allows for reconstructing objects made of multiple materials. Under this assumption, searching for l_P can be posed as a energy minimization that measures the unsmoothness of l_P. Specifically we use the term

$$D_p = var(\frac{\delta}{\delta c} f(l_P^{\lambda})) \tag{10}$$

where $var(\cdot)$ denotes the statistical variance and f is the line in the epipolar volume which is defined by the pixel position and depth label. Results for our GPU/CPU implementation of the described energy function are presented in Section 6.

5 Algorithm and Implementation

To implement the energy minimization we propose a modification of the α-expansion multi label graph cut method introduced in [2]. In order to use this optimization algorithm the resulting depth values have to be discretized. Let L define the set containing the discrete labels. The cardinality of L determines the quality of the computed depth image. Thus to improve results a finer discretisation can be used at the cost of additional computation time.

The α-expansion graph cut algorithm is divided in several steps. First for each depth label a graph is built, encoding both pixel neighborhood smoothness and pixel depth fit. Then in each iteration the minimal cut on the graph is computed using the MAXFLOW algorithm [20].

Specifically, let $G = (V, E)$ be a graph constructed for the energy minimization problem and m be the mapping

which maps pixels to labels. For each label $\alpha \in L$ compute in each iteration the energy function

$$\hat{m} = \arg\min E(m') \tag{11}$$

where m' is the mapping obtained after the α-expansion step on a graph G by using the current labeling function m. The energy function to minimize has the following form:

$$E(f) = \sum_{(p,q)\in N} V_{p,q}(m_p, m_q) + \sum_{p\in P} D_p(l_P) \tag{12}$$

where N is the set of interacting pairs of pixels. Typically the neighborhood N is built from adjacent pixels, but it can be arbitrary. D is the data term introduced in Section 4.3. The term V regularizes the neighborhood and is discussed in more detail in Section 5.1.

Looking at D reveals that it can be computed independently for each pixel. To accelerate it we shifted the computation on the GPU. Modern graphics hardware allows us to reprogram its pipeline so we can use it for our own purpose. The GPU's fragment shader operates in parallel on each pixel and is programmed to compute the D_p. Since the graph has always got a constant number of edges e_p with weight $D_p(l)$, we can compute these weights on the GPU for all edges e_p. Utilizing the GPU reduces the runtime to a third in comparison to the time needed for a CPU only implementation (see Table 1).

Table 1: Runtimes of the algorithm executed on Intel Pentium 4 with 2.4GHz 512KB Cache and 2GB RAM. The GPU is a Nvidia GeForce 6800 GT (NV45) graphic chip. The table shows time needed to accomplish one graph cut cycle on the whole image volume of the elephant scene for 16 labels.

Resolution	*dT CPU*	*dT GPU*
128x128x10	55 sec	22 sec
256x256x10	217 sec	71 sec
512x512x10	749 sec	233 sec

5.1 Neighborhood

Assuming the surface is piecewise smooth makes the optimization algorithm more robust against outliers and finds information in parts of the surface that have little or no texture information. To account for this we introduce a neighboring term based on the Pott's Model as proposed in [2]. Because of complexity reasons, the neighborhood is restricted to each camera view but could be extended to neighboring views as well. For the neighborhood term V, computed for each pair of neighboring pixels $(p, q) \in N$, we used the following term

$$V_{p,q}(m_p, m_q) = \begin{cases} 0 & |m_p - m_q| < d_1 \\ 2K & d_1 \leq |m_p - m_q| < d_2 \\ K & \text{else} \end{cases} \tag{13}$$

The constant K determines the strength of the neighborhood regularization. Constants d_1, d_2 control the smoothness of the result and the handling of depth discontinuities.

5.2 Occlusion

We implement the α-expansion graph cut algorithm to find the dense depth maps by global energy minimization similar to [2]. The difference in the optimization however is the computation of the data term in the energy function. Unlike previous methods, we do not use a separate term for visibility but handle occlusion directly in the data term. This has the advantage that partly occluded pixels are not omitted from reconstruction. To achieve this, we extend the α-expansion graph cut algorithm in the following way.

- Initialize the reconstruction volume to λ_{max}
- Solve the α-expansion iteration for the ordered labels from λ_{max} to λ_{min} for all cameras. Thus in each step the next depth value is probed for the complete volume.
- In each α step extract for the current λ_α and each p the line l_P
- Invalidate pixels on l_P with $\lambda > \lambda_\alpha + M$ to get $\hat{l}_p$
- Computing $D_P(\hat{l}_p)$ then readily includes occlusion handling as discussed in Section 4.2

The constant M is introduced for robustness reasons and is set in relation to the cardinality of the labelset L, e.g. $M = \lceil \log \lambda_{max} \rceil$. The iteration continues until $Iter_{max}$ is reached or no change in the energy occurs during a complete α cycle.

In each iteration we compute the α-expansion step for all cameras. Since the initial value of each pixel is set to the closest possible position, there is no occlusion culling for pixels in the first M iterations. Under the assumption of decreasing energy as the true disparity is approached (see Figure 5), we conclude most pixels that have reached their lowest possible energy are truly at the corresponding depth in the scene. Thus it is plausible to remove pixels from l_P that have higher label values during the reconstruction process to form the $\hat{l}_P$. Although there are pixels violating this assumption, later iterations are probable to resolve the errors introduced.

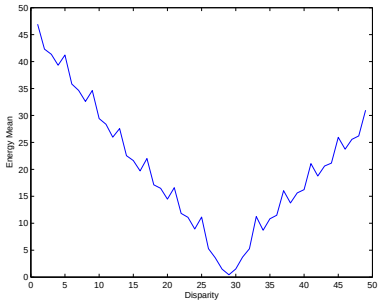

Figure 5: The plot shows the mean energy distribution for the pixels of the elephant scene that have their minimum at disparity $\lambda = 29$.

For the examples presented in this work, we found that this assumption is strong enough to reconstruct objects with occlusion in the proposed way. The validity of the assumption about the energy functions is further underlined by the observed decreasing number of labels changed in each iteration. All examples needed no more than three cycles to find the global energy minimum.

6 Results

Our test set consists of synthetic objects raytraced with POVRay [21] and ground truth. The scenes are illuminated with multiple light sources to simulate a complex lighting setup. No information about the lighting setup was incorporated in the reconstruction results. We applied three representative non-lambertian materials to the objects to compare performance to the well-known graph cut algorithm introduced in [3]. The materials include plastic, chrome and an anisotropically reflecting brushed metal. Results were computed on image sizes of 512×512 from 10 views. The cameras were equally spaced on the same baseline perpendicular to the viewing direction.

Figure 6 shows the elephant model rendered with a chrome material. We used 16 labels for the depth reconstruction. The results obtained with our novel approach give improved results over the standard color constancy assumption. Most of the artifacts found with the constant color energy term have vanished. Other result on reconstructing shiny surfaces like the plastic Happy Buddha model [4] are shown in Figure 1 and the Dragon model with a anisotropically reflecting brushed metal surface in Figure 7. For the Buddha model we used 16 and for the dragon model 32 labels for reconstruction. Again the results improve considerably if our proposed smoothness assumption defined on lines in the epipolar volume is used for the reconstruction.

7 Discussion and Future Work

The main contribution of this work is the integration of prior knowledge about the surface reflection properties in the reconstruction process. We showed how to enhance the existing graph cut reconstruction method using Epipolar Volumes to perform multi view reconstruction utilizing this information. Improved reconstruction results in comparison to standard color constancy methods have been obtained for shiny surfaces. The proposed method is also fit to handle more complex scenes when more information, e.g. the environment map, the BRDF or the lighting setup is known. Using both CPU and GPU in the implementation of our algorithm results in accelerated reconstruction runtimes. In comparison to a CPU only implementation we achieve a speedup factor of three. Since we use graph cuts to optimize the energy function we have to discretize the depth. Thus the quality of the reconstructed volume depends on a sufficient discretization of the result space.

In future work, we like to explore more ways to incorporate prior knowledge like the light setup or the environment map. First tests to reconstruct perfect mirroring surfaces have already shown promising results. For this reconstruction problem, the normals of the surface must also be estimated which causes problems when using discrete optimization methods. Future research on other continuous optimization methods to solve this problem are an interesting application of the proposed framework.

References

[1] F. E. Nicodemu, J. C. Richmond, J. J. Hsia, I. W. Ginsber, and T. Limperis, *Geometrical Considerations and Nomenclature for Reflectance*, U. S. Dept. of Commerce, 1977.

[2] O. Veksler Y. Boykov and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," in *IEEE Transactions on pattern analysis and machine intelligence*, 2001, pp. 1222–1239.

[3] V. Kolmogorov and R. Zabih, "Multi-camera Scene Reconstruction via Graph Cuts," in *European Conference on Computer Vision(ECCV)*, 2002, pp. 82–96.

[4] "The Stanford 3D Scanning Repository," http://graphics.stanford.edu/data/3Dscanrep.

[5] B. Goldluecke and M. Magnor, "Joint 3D-Reconstruction and Background Separation in Multiple Views using Graph Cuts," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003, pp. 683–694.

[6] L. Wang J. Davis, R. Yang, "BRDF Invariant Stereo using Light Transport Consistancy," in *International Conference on Computer Vision(ICCV)*, 2005, pp. 436–443.

[7] D. Bhat and S Nayar, "Stereo in the Presence of Specular Reflection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1995, pp. 1086–1092.

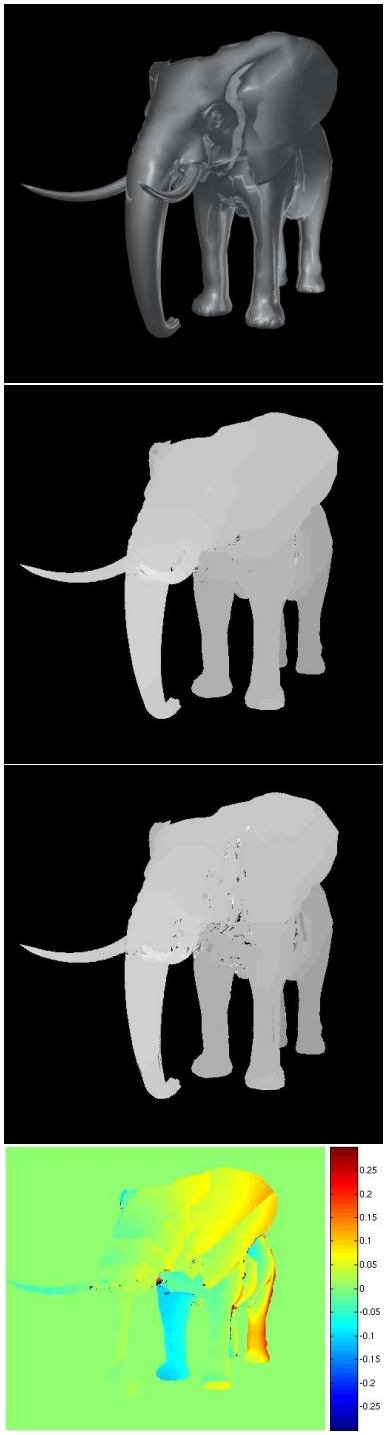

Figure 6: Results of the elephant model seen from camera 5. 10 different views are used for reconstruction. We applied a chrome material to the object. The second image shows the reconstructed depth map found with our method. For comparison the next image shows the results obtained with the color constancy energy term. The last image shows the differences of our results compared to ground truth measured in percentage of the label depth range. The error increases with the depth since resolution decreases with increasing depth.

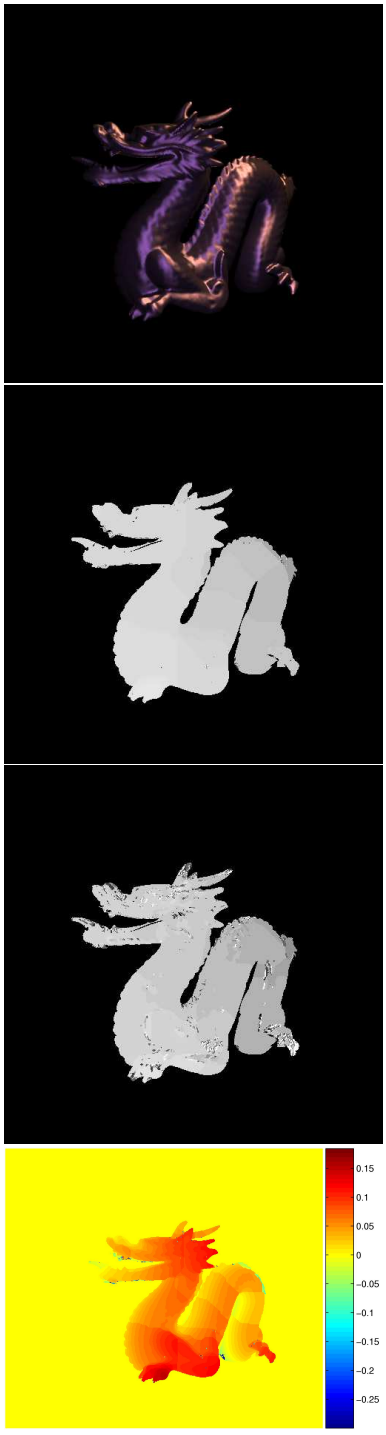

Figure 7: Results of the Dragon Model [4] with an anisotropically reflecting brushed metal material applied as seen from camera 5. 10 different views are used for reconstruction. The second image shows the depth map reconstructed with our proposed algorithm. The third image shows the results obtained using the color constancy energy term. Large errors due to the violation of the color constancy assumption are found. The last image shows the difference between our results and ground truth measured in percentage of the label depth range.

[8] Y. Li, S. Lin, H. Lu, S. B. Kang, and H.-Y. Shum, "Multi-baseline Stereo in the Presence of Specular Reflections," in *International Conference on Pattern Recognition (ICPR)*, 2002.

[9] P. Gargallo and P.F. Sturm, "Bayesian 3D Modeling form Images Using Multiple Depth Maps," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 885–891.

[10] R. L. Carceroni and N. K. Kutulakos, "Multi-View Scene Capture by Surfel Sampling," in *International Conference on Computer Vision(ICCV)*, 2001, pp. 60–67.

[11] K. N. Kutulakos and S. M. Seitz, "A Theory of Shape by Space Carving," in *International Conference on Computer Vision(ICCV)*, 1999, pp. 307–314.

[12] G. Vogiatzis, P. H. Torr, and R. Cipolla, "Multi-View Stereo via Volumetric Graph-Cuts," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 391–398.

[13] R. Yang, M. Pollefeys, and G. Walch, "Dealing with textureless regions and specular highlights - a progressive space carving scheme using a novel photo-consistency measure," in *International Conference on Computer Vision(ICCV)*, 2003, pp. 576–584.

[14] T. Yu, N. Xu, and N. Ahuja, "Shape and View Independent Reflectance Map from Multiple Views," in *European Conference on Computer Vision(ECCV)*, 2004.

[15] H. Jin, S. Soatto, and A. Yezzi, "Multi-View Stereo Beyond Lambert," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003, pp. 171–178.

[16] R. Bolles, H. Baker, and D. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, pp. 7–55, Mar 1987.

[17] A. Criminisi, S.Kang, R. Swaminathan, R. Szeliski, and P. Anandan, "Extracting Layers and Analyzing their Specular Properties Using Epipolar-Plane-Image Analysis," in *Computer Vision and Image Understanding (CVIU)*, 2005, pp. 51–85.

[18] E. H. Adelson and J. R. Bergen, *Computational Models of Visual Processing*, chapter The Plenoptic Function and the Elements of Early Vision, pp. 3–20, MIT Press, 1991.

[19] R. Hartley and H. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[20] Y. Boykov and V. Kolmogorov, "The MAXFLOW algorithm," http://www.cs.cornell.edu/People/vnk/software.html.

[21] Persistence of Vision Pty. Ltd. (2004), "Persistence of Vision (TM) Raytracer," http://www.povray.org/.