Estimating Planar Patches for Light Field Reconstruction

Andrew Mullins, Adam Bowen, Roland Wilson, Nasir Rajpoot
Signal & Image Processing Group
Department of Computer Science
University of Warwick
Coventry, CV4 7AL, England
{andy, fade, rgw, nasir}@dcs.warwick.ac.uk

Abstract

Light fields are known for their potential in generating reconstructions of a scene from novel viewpoints without need for a model of the scene. Reconstruction of novel views, however, often leads to ghosting artefacts, which can be relieved by correcting for the depth of objects within the scene using disparity compensation. Unfortunately, reconstructions from the disparity information suffer from a lack of information on the orientation and smoothness of the underlying surfaces. In this paper, we propose a novel representation of the surfaces present within a scene using a planar patch approach. We discuss the process of estimating patches and introduce a reconstruction algorithm designed to exploit this patch information to produce visually superior reconstructions at higher resolutions. Experimental results demonstrate the effectiveness of this reconstruction technique when compared to traditional reconstruction methods.

1 Introduction

A Light Field [7] captures a large array of images of a scene in a representation that allows fast reconstruction from an arbitrary location and preserves view dependent effects. The scene is represented as a number of camera viewpoints of a common imaging plane. The pixel samples then correspond to the intersections of a ray with the image plane and the camera plane. Traditional light field reconstruction algorithms exploit this efficient data structure to rapidly sample light rays for every pixel being reconstructed. Unfortunately, it is often impractical or even impossible to capture the camera plane at sufficient resolution to represent all the desired viewpoints, resulting in noticeable artefacts in the reconstructions. Attempts have been made to alleviate this problem using variable focus and aperture [5], compensation with a low resolution model [4] and image warping [9]. Other techniques for image based rendering can also be applied to light field data, such as space carving [8] and photo-consistency approaches [3].

In fact, there is significantly more information in a light field than is exploited by traditional reconstruction approaches. Traditional reconstruction does not take advantage of the fact that all the camera views are of the same object to infer properties of the object.

By examining the light field data we can obtain information about the object of interest that will allow us to improve our reconstructions. Typically, this is the approach taken in image warping [9]. Warping extracts disparity information from the available images to then warp them to the novel viewpoint. However, this introduces problems during reconstruction, most significantly dealing with multiple conflicting samples of the same pixel and filling 'holes' in the reconstructed image. These problems arise because disparity information between images is not sufficient to model the shape and orientation of the surfaces present in the scene and so occlusion boundaries cannot be properly reconstructed. Other methods for computing reconstructions from light fields include photo-consistency [3] and space carving [8]. Using a photo-consistency approach for reconstruction is very slow, as not much preprocessing can be performed, whilst using a space carving approach discards the view-dependent information.

We present a novel representation of the surfaces present in the scene using planar patches, and an algorithm for the reconstruction of these patches when the patch estimates may be unreliable. In section 2 we introduce a locally planar representation of light field data, and discuss how the patches are estimated. We then describe our reconstruction algorithm in section 3. Section 4 summarises our results thus far, and finally section 5 presents our conclusions and further avenues of research.

2 Planar patch estimation

Given a set of light field images, our aim is to construct a model of the scene captured by those images. Previous work has attempted to estimate scene geometry by representing it as a set of depth values [10], voxels [11], or a polygon mesh [6]. Our aim is to divide each image into a set of $N \times N$ square blocks, and treat each block as the perspective projection of some planar quadrilateral P in 3D space which is at a distance z from the camera, and whose normal is rotated through θ, ϕ relative to the x and y axes (respectively). Typically, the images which constitute a light field are pre-captured or pre-rendered from locations which lie on a regularly spaced grid (see figure 1) in 3D space. Thus our goal is to estimate, for each camera located at u, v on this grid, and for each block (centred at pixel s,t) within the images obtained at these points, a patch

$$P_{uvst} = (z_{uvst}, \theta_{uvst}, \phi_{uvst}). \tag{1}$$

Any camera u, v (excluding those which are at the edges of the grid) has a set of 8 neighbouring cameras N_{uv}. We attempt to recover the patches P_{uvst} by minimizing the function

$$E_{uvst}(z,\theta,\phi) = \sum_{x=s-N/2}^{s+N/2} \sum_{y=t-N/2}^{t+N/2} \sum_{pq \in N_{uv}} (I_{uv}(x,y) - I_{pq}(r_{uvpq}(x,y,z,\theta,\phi,s,t))^2, \tag{2}$$

where $I_{ab}(c,d)$ is the image function for the camera a,b and pixel c,d, and where $r_{uvpq}(x,y,z,\theta,\phi,s,t)$ is the re-projection of pixel x,y in camera u,v into camera p,q, given that it belongs to a patch with parameters, z,θ,ϕ and whose centre projects to the point s,t in the original camera.

Unfortunately, our search space is now 3 dimensional, and so an exhaustive search is intractable. Our approach is to use simulated annealing to minimize the above function, drawing our starting points $P = (z,\theta,\phi)$ from

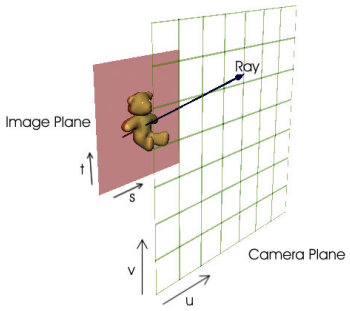


Figure 1: The cameras are located on a regularly spaced grid, defined by the co-ordinates u, v and capture an image function $I(s,t)$.

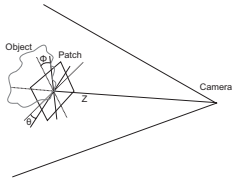


Figure 2: The parametrisation of our patches $P = (z, \theta, \phi)$.

$$z \sim N(f, \sigma_1), \tag{3}$$

$$\theta \sim N(0, \sigma_2), \tag{4}$$

$$\phi \sim N(0, \sigma_3), \tag{5}$$

where f is the distance to the focal plane and σ_1 is proportional to the depth of field. The variances σ_2 and σ_3 are determined empirically.

3 Reconstruction Algorithm

The estimation of planar patches, as described in section 2, takes place for every camera. The generated patches are locally consistent with the viewpoint from which they were estimated. If our patch data were perfect, this would be sufficient to construct a model of the object and recreate the novel view using traditional rendering techniques. However, the patches are estimated from images which are subject to noise and non-Lambertian surface properties and are therefore prone to error. Our reconstruction algorithm takes account of these potential discrepancies by dividing the process into two stages. During the first stage a reconstruction is generated for every camera independently, using the patch data for that camera alone. The second stage then looks at the the consistency of

the data across all the reconstructions to eliminate erroneous patches and select the best reconstruction. Figure 3 shows how the reconstruction algorithm proceeds.

3.1 Independent Reconstruction

Each patch is estimated using a block in the source camera's image. We generate an individual camera's estimate of the reconstruction by calculating a quadrilateral in 3D space that corresponds to the image block used to generate each patch, as illustrated by figure 2. Figure 4(a) shows patches for one camera in the 'Teddy' light field. The 'holes' seen in the image are regions that the camera cannot see, and so has no patch information for - most notably a 'shadow' of teddy is clearly visible on the background. Once the quadrilaterals have been computed, they are then textured and projected into the virtual viewpoint where a depth test is applied. Figure 4(b) shows the result of texturing and rendering the patches seen in figure 4(a) using standard OpenGL methods. We obtain an image similar to this for every available viewpoint. Only nearest neighbour interpolation is applied to the textures at this stage, to avoid blurring the textures during the second stage. This independent reconstruction stage can use graphics hardware to render the quadrilaterals as polygons and so is very fast.

3.2 Combining Reconstruction Images

Once each camera has generated an estimate of the reconstructed image, we attempt to identify the surfaces that are present at each reconstruction pixel using a clustering approach. For every pixel we wish to reconstruct, we have a colour sample and depth available from the estimate generated by each camera. Clustering these four dimensional vectors (red, green, blue and depth) gives us an estimate of the surfaces present in the reconstruction, and their corresponding depths.

To obtain the surface estimates, we apply a hierarchical clustering algorithm that finds the minimum number of clusters such that the total squared error within the cluster is below a threshold value. In our experiments we have found that, when the colour and depth values are between 0 and 1, a threshold between 0.1 and 0.3 gives good clustering of the surfaces. The result is a variable number of clusters for each pixel that estimate the surfaces present along the ray. Small clusters may correspond to erroneous patches whilst larger clusters may correspond to genuine surfaces.

Given these clusters and their corresponding depths, we wish to select the cluster most likely to provide an accurate reconstruction. In other words, we wish to maximise

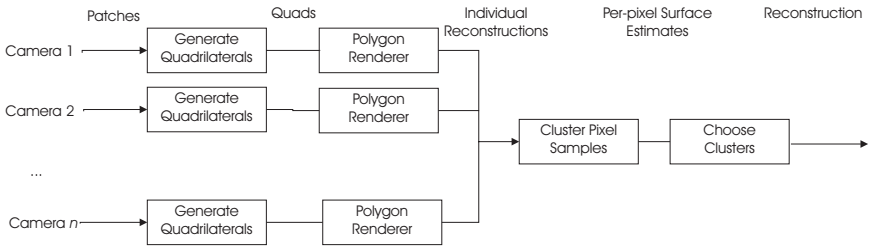


Figure 3: Reconstruction Algorithm

(a) Shaded (b) Textured

Figure 4: Surface patches estimated from scene geometry for a single camera

the conditional probability

$$P(c_i|c_1,c_2...c_n) \tag{6}$$

for the selected cluster c_i and sample clusters $c_1,c_2...c_n$. Bayes' law gives us

$$P(c_i|c_1,c_2...c_n) = \frac{P(c_1,c_2...c_n|c_i).P(c_i)}{P(c_1,c_2...c_n)}. \tag{7}$$

Since $P(c_1,c_2...c_n)$ is constant across our maximisation, it can be ignored. This simplifies the problem to maximising

$$P(c_1,c_2...c_n|c_i).P(c_i) \tag{8}$$

$P(c_i)$ is some measure of how reliable our cluster is. There are two factors to consider when calculating this measure. Firstly, we must consider the number of cameras that support the hypothesis that this cluster is a valid surface in our scene. Secondly, we must consider how much we trust the information provided by the supporting cameras. To achieve this, we assign to each camera j a weight w_j, the weight is computed as the dot (scalar) product of the direction of camera j and the direction of the reconstruction camera. If the direction of camera j is given by d_j and the direction of the reconstruction camera is d_{camera} then we find the weight as

$$w_j = \text{clamp}(0,(d_k.d_{\text{camera}})^\rho,1) \tag{9}$$

where ρ is a tuning parameter used to control how closely aligned cameras must be before they are trusted and the clamp function clamps the value to the range $[0,1]$. Typically values of 5-8 cut out undesirable viewpoints. We define the probability of the cluster as

$$P(c_i) = \frac{\sum_{j\in c_i} w_j}{\sum_{k=1}^{C} w_k} \tag{10}$$

where $j \in c_i$ if camera j is in cluster c_i and C is the total number of cameras.

We now need to decide how consistent the surfaces are with the selected surface. We say a surface is consistent with another surface if it occludes that surface, hence

$$P(c_1,c_2...c_n|c_i) = \frac{\sum_{j=1}^{n} \text{occludes}(c_i,c_j)}{n} \tag{11}$$

Reconstruction Algorithm	PSNR	Time Complexity
Traditional Reconstruction	24.5dB	$O(N)$
Warping (perfect disparity maps)	31.5dB	$O(N)$
Warping (estimated disparity maps)	28.4dB	$O(N)$
Photo-consistency	27.0dB	$O(N.D.C^2)$
Patch Rendering (from geometry)	32.0dB	$O(N.C^2)$
Patch Rendering (from estimates)	32.6dB	$O(N.C^2)$

Table 1: Comparison of reconstruction algorithms in terms of PSNRs and algorithmic complexity.

where

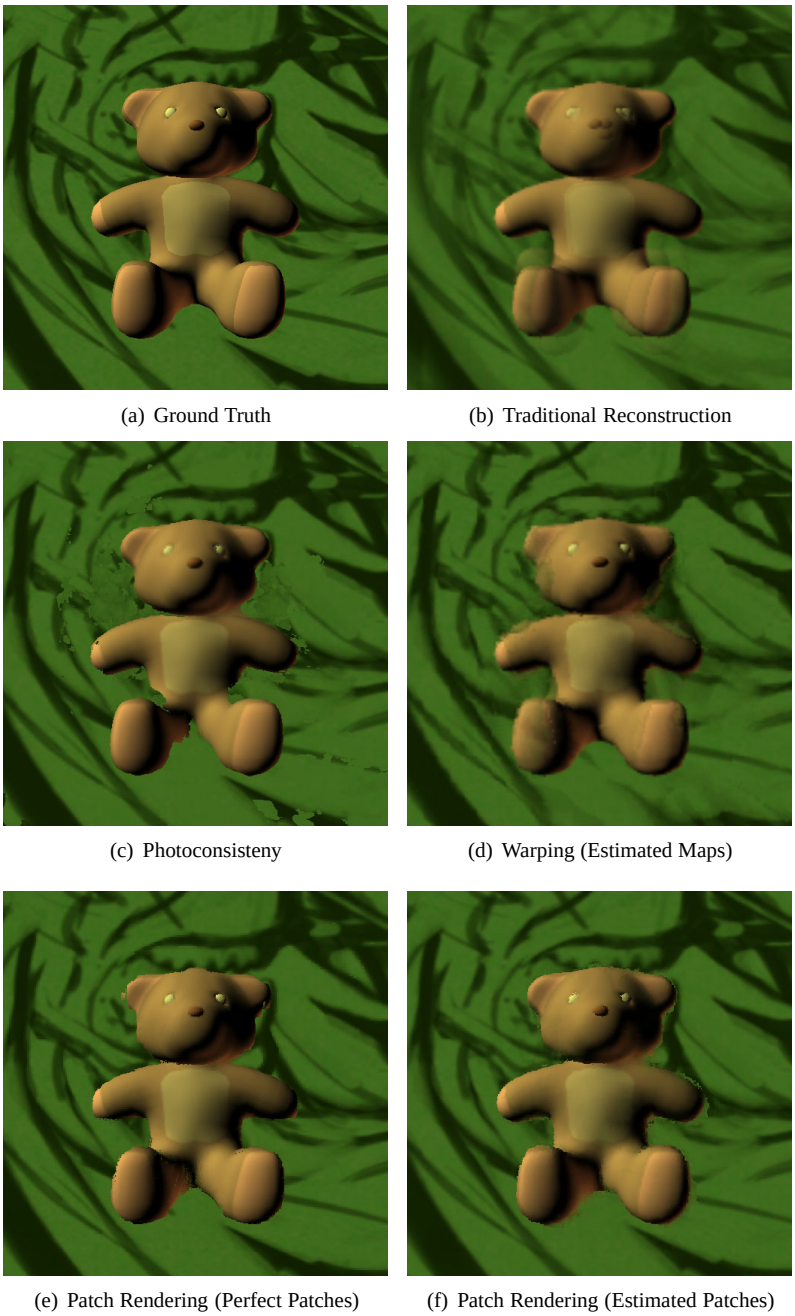
$$\text{occludes}(c_i, c_j) = \begin{cases} 1 & z_i \leq z_j, \\ 0 & \text{else.} \end{cases} \tag{12}$$

and z_i is the depth of the centroid of cluster c_i. Combining these two probabilities as in equation 8 gives us a measure of the quality of the surface represented by cluster c_i which we can then maximise for a value of c_i.

4 Results

We compared results of reconstruction using our patch model based rendering with three other techniques: traditional reconstruction, warping [9], and photo-consistency based reconstruction [3]. Our synthetic light field consists of 64 images of 'teddy' (in an 8×8 arrangement), and contains significant depth variation. Each source image was a 256×256 pixel image. In order to assess the quality of different reconstructions, we computed the peak-signal-to-noise-ratio (PSNR) of the reconstructed images for all viewpoints as compared to the ground truth reconstruction. The reconstruction PSNR and time complexity for all the algorithms are summarised in Table 1, where N is the number of pixels, C is the number of cameras, and D is the number of depth samples (for the photo-consistency approach). In the case of the photo-consistency reconstruction, we maximised the photo-consistency metric described in [1]. Whilst the PSNRs are comparable, the patch based algorithm produces noticeably sharper and higher quality reconstructions.

Figure 5(b) shows the ghosting and blurring artefacts that typically result from a light field reconstruction when the camera plane is heavily under-sampled. Figures 5(c) and 5(d) alleviate the problems with the traditional reconstruction approach by realigning the images used in the reconstruction using disparity information. Reconstruction from perfect disparity maps suffers from hole filling problems due to occlusion between the legs and under the arm. This is because the warping approach only considers at most the 4 closest cameras for each pixel and in this case none of the cameras can see the desired region. It also suffers problems across the front of the legs. Because it has no model of how smooth or disjoint the surface is it cannot correctly interpolate nearby samples that belong to the same surface, the result is that parts of the background 'show through' the legs when no sample on the leg warps to the pixel. These problems are not visible when using the estimated maps because the error in the maps prevents the samples from aligning. However, the lack of accuracy shows through when the samples from contributing



(a) Ground Truth

(b) Traditional Reconstruction

(c) Photoconsisteny

(d) Warping (Estimated Maps)

(e) Patch Rendering (Perfect Patches)

(f) Patch Rendering (Estimated Patches)

Figure 5: Reconstruction Results for a Camera

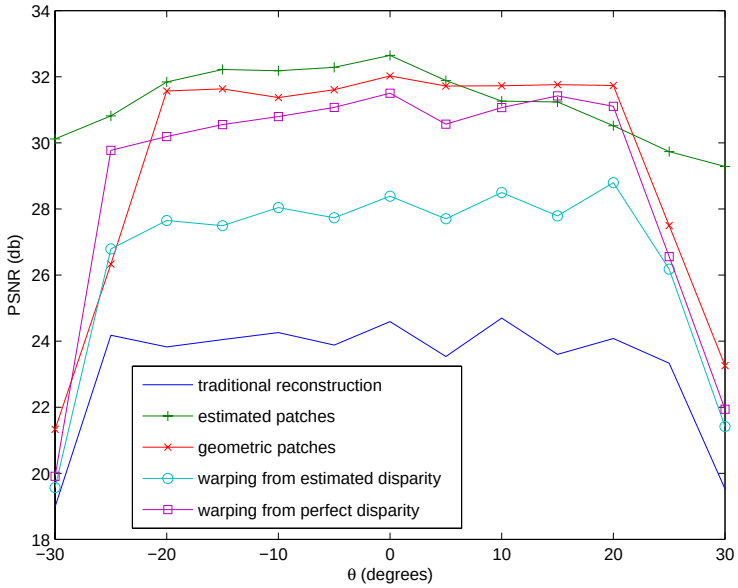


Figure 6: Reconstruction quality (PSNR in dB) as we pan horizontally across the light field

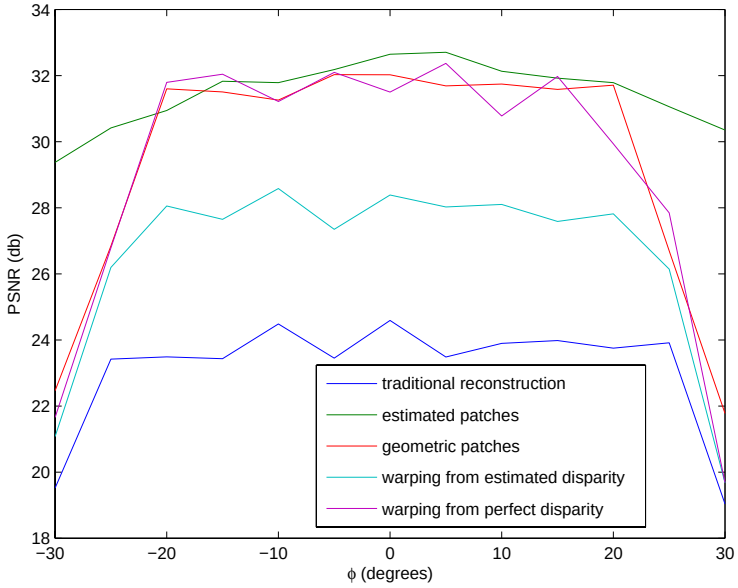


Figure 7: Reconstruction quality (PSNR in dB) as we pan vertically across the light field

(a) Perfect (b) Estimated

Figure 8: Shaded images of some of the patches used for our reconstructions

cameras are blended. Blending samples that do not come from the same surface results in a loss of detail in the image and often undesirable blurring or ghosting in the reconstruction. Figure 5(e) shows the reconstruction using perfect patches. This reconstruction is visually significantly superior to the other methods shown, due to the accurate recovery of the edges. Because these reconstructions are generated at twice the resolution of the original light field, the technique is effectively achieving super-sampled reconstruction - making it more suitable for reconstructing scenes at different resolutions and from closer camera positions. The notable artefacts occur where part of the ear has been lost due to few cameras providing a reliable patch and a number of single pixel errors which could easily be restored using a prior-based refinement of the reconstruction. Figure 5(f) shows the reconstruction from estimated patches. Whilst not as visually pleasing as the results generated using perfect patches the estimated patches produce reconstruction results on a par with those using scene geometry - demonstrating the combined effectiveness of our patch estimation and reconstruction algorithms. Figure 8 compares some of the patch estimates with the perfect estimates, illustrating the quality of our patches, and a few erroneous patches that are eliminated by the clustering stage of the reconstruction algorithm. Figure 6 shows how the reconstruction PSNR varies as we pan horizontally around the light field and figure 7 as we pan vertically around the light field. The peaks correspond to regions where the viewpoint aligns more closely with the source viewpoints. There are significant drops in PSNR towards the extreme angles because the arrangement is such that no camera can see some of the background needed to create the reconstruction.

5 Conclusions

We have presented a novel method of estimating geometry found in light field data sets. Using reconstruction techniques previously shown in [2] we have demonstrated the combined effectiveness of patch estimation and reconstruction. The traditional and warping reconstruction approaches are computationally efficient, but do not exploit all the information that can be extracted from the data set to produce the highest quality reconstructions. Instead they rely on a high volume of data to create accurate and high quality reconstructions - which is not ideal when it comes to the coding and transmission of light field data sets. Although our method is more computationally demanding, it is still relatively simple and scalable to higher resolutions. It provides more information on the structure of a scene whilst retaining the view-dependent properties of the surfaces in the scene. We can also generate visually superior reconstructions utilising the inherent super-resolution information available in light field data sets.

Acknowledgements

This research is funded by EPSRC project 'Virtual Eyes', grant number GR/S97934/01.

References

[1] A. Bowen, A. Mullins, N. Rajpoot, and R. Wilson. Photo-consistency and multiresolution based methods for light field disparity estimation. In *Proc. VIE 2005, Glasgow, Scotland, April 2005*, April 2005.

[2] A. Bowen, A. Mullins, R. Wilson, and N. Rajpoot. Light field reconstruction using a planar patch model. In Heikki Kälviäinen, Jussi Parkkinen, and Arto Kaarna, editors, *To Appear In: Proc. 14th Scandinavian Conference on Image Analysis (SCIA 2005)*. Springer-Verlag, June 2005.

[3] A.W. Fitzgibbon, Y. Wexler, and A. Zisserman. Image-based rendering using image-based priors. In *Ninth IEEE International Conference on Computer Vision, Nice, France, October 2003*, volume 2, pages 1176–1183, October 2003.

[4] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proceedings of ACM SIGGRAPH '96, New Orleans, LA, August 1996*, pages 43–54. ACM Press, New York, August 1996.

[5] A. Isaksen, L. McMillan, and S. J. Gortler. Dynamically reparameterized light fields. In Kurt Akeley, editor, *Proceedings of ACM SIGGRAPH 2000, New Orleans, Louisiana, July 2000*, pages 297–306. ACM Press, New York, July 2000.

[6] J. Isidoro and S. Sclaroff. Stochastic mesh-based multiview reconstruction. In *The First International Symposium on 3D Data Processing Visualization and Transmission*, July 2003.

[7] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of ACM SIGGRAPH '96, New Orleans, LA, August 1996*, pages 31–42. ACM Press, New York, August 1996.

[8] W. Matusik. Image-based visual hulls. Master of science in computer science and engineering, Massachusetts Institute of Technology, February 2001.

[9] H. Schirmacher. Warping techniques for light fields. In *Proc. Grafiktag 2000, Berlin, Germany, September 2000*, September 2000.

[10] I. O. Sebe, P. Ramanathan, and B. Girod. Multi-view geometry estimation for light field compression,. In *Proc. Vision, Modeling, and Visualization VMV-2002, Erlangen, Germany*, pages 265–272, November 2002.

[11] S. M. Seitz and C. R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference, Puerto Rico, June 1997*, pages 1067–1073, June 1997.