Scene Reconstruction by Greedy Belief Propagation

Chris J. Forne, Michael P. Hayes
Department of Electrical and Electronic Engineering,
University of Canterbury, Private Bag 4800, Christchurch, New Zealand
c.forne@elec.canterbury.ac.nz

Abstract

This paper presents a new hybrid algorithm for reconstructing three-dimensional scenes from multiple camera images. Based on both belief propagation and the greedy technique, the new algorithm combines the two approaches overcoming some of the problems specific to each. The resulting algorithm is essentially a dynamic form of belief propagation, where the local compatibility functions vary from iteration to iteration under the guidance of a greedy approach. This enables detailed statistical prior information to be incorporated into the reconstruction through the belief propagation process, while at the same time dealing with occlusions through the greedy technique. To demonstrate the advantages of this approach, the algorithm was tested on several simulated data sets. Results show an improvement in performance over both the original greedy and belief propagation algorithms, although execution time is slightly longer.

Keywords: stereo vision, multiple camera, belief propagation, greedy algorithm

1 Introduction

The problem of reconstructing or estimating the structure of a three-dimensional scene from several photometric images is fundamental to many aspects of image and vision computing. The problem was first investigated in 1849, when Aim Laussedat used terrestrial photographs for topographic map compilation. At about the same time, investigations into human or biological stereopsis began, following the invention of the stereoscope by Sir Charles Wheatstone. Since then a large amount of research has been undertaken, including recent developments within the fields of machine vision and computer graphics.

Such a diverse background has lead to a wide range of ideas and approaches for performing stereo reconstruction. Initial techniques were based on image matching [1], where points, regions or features were matched between pairs of images. The depth of each matched primitive was then found by simple triangulation. This worked tolerably well for simple scenes but did not make good use of the available information. In recent years, research has centered around applications in machine vision and computer graphics, where accurate and detailed models of often complex scenes are required. For such applications, traditional techniques are unsuitable, requiring new and improved approaches to deal with multiple cameras, multiple surfaces, and widely varying views. This has lead to various global optimisation techniques [2, 3, 4], as well as to several volumetric algorithms, where the scene is reconstructed directly in a true three-dimensional space [5, 6, 7]. Unfortunately, little work has been done on algorithms that combine these two approaches.

To address this, a new hybrid algorithm based on a Bayesian approach is presented, that performs global optimisation within a volumetric framework. This offers the benefits of a volumetric approach, where the interaction between scene parameters can be accurately modelled in three-dimensions, while still providing efficient global optimisation over the entire solution space. Based on a combination of Belief Propagation and the Greedy technique, the proposed algorithm successfully combines the two approaches.

First a review of related work is given, including a discussion of Belief Propagation and the Greedy approach. The problem of scene reconstruction is defined more precisely in Section 3, along with the system model that will be used. Section 4 describes the proposed Greedy Belief Propagation algorithm in detail and explains the theory behind it. Experimental results are then shown in Section 5, demonstrating the effectiveness of this algorithm on several test scenes, followed by a discussion of the algorithm and results in Section 6. Finally a brief conclusion of the work is given in Section 7.

2 Related Work

Recent research in scene reconstruction has focused on the estimation of complex scenes from multiple camera images. Such systems, often containing numerous surfaces, discontinuities and occlusions, introduce a number of problems that are not easily dealt with using traditional stereo matching approaches.

2.1 Global Optimisation

To address some of these issues and improve the scene reconstruction process, a lot of recent work has focused on global optimisation techniques [2, 3, 4]. These are techniques which attempt to find the overall optimal solution to a problem, rather than simply some local solution. Of these, the best results to date have been obtained though the use of Graph Cuts [3] and Belief Propagation [2]. For a comprehensive review of these, and other recent stereo algorithms, applied to a two camera system, the reader is referred to an excellent paper by Scharstein and Szeliski [8]. In particular Belief Propagation has been shown to be an effective and highly parallel algorithm that can be applied to a wide range of inference problems.

2.2 Belief Propagation

"Belief Propagation", proposed by Pearl [9] is one of several closely related message passing algorithms that have been independently developed for solving inference problems on probabilistic models. Equivalent or very closely related algorithms include the Viterbi algorithm, the turbo-decoding algorithm, and the Kalman filter. All of these algorithms are designed to either maximise the joint posterior probability distribution of the entire graph or determine the marginal posterior probability distribution of individual nodes. What is sometimes confusing, is that there are two forms of the belief propagation algorithm, one for each of these situations. These are referred to as the "max-product" and "sum-product" algorithms respectively. In either case, messages are iteratively sent between neighbouring nodes in the graphical model until a final solution is obtained. The algorithms are exact when the graphical model has a tree structure but only approximate when the graph contains cycles. Fortunately, surprisingly good results are often obtained even in the presence of loops.

Sun et al [2] recently applied belief propagation to the scene reconstruction problem, with some excellent results. In their approach the scene is represented as a depth map relative to one of the images. The conditional likelihood of each point given the image data is then modelled using additional observation nodes and associated compatibility functions.

As with the majority of approaches to scene reconstruction, one of the main problems with their method is that the interaction between scene points is poorly modelled. In particular they assume occlusions are statistically independent of the estimated depth map. This is a rather poor assumption and leads to an inaccurate set of compatibility functions. This in turn can cause the belief propagation process to converge to an undesirable solution, resulting in a reduced quality of the scene estimate.

2.3 Volumetric Methods

To overcome some of the problems associated with occlusions and improve results, a number of scene or volumetric based methods have been proposed [5, 6, 7]. These are especially helpful when dealing with complex surfaces or general multiple camera systems. In most cases the scene is represented using a true three-dimensional model, rather than as a depth-map or a single surface. This is preferable to the traditional representations, as the interaction between scene parameters can be accurately modelled in three-dimensions. A recent survey of volumetric methods is given in [10]. The problem with most of these approaches, however, is that the global optimisation techniques used are rather poor, resulting in suboptimal scene reconstructions.

2.4 The Greedy Approach

In an attempt to improve the results of volumetric scene reconstruction, Forne and Hayes proposed a greedy approach [11], which attempts to find the Maximum A Posteriori (MAP) estimate of the scene. Beginning with a transparent estimate, points are progressively assigned as opaque until a complete scene estimate has been formed. This is done by selecting the most likely surface point at each iteration, and updating the remaining visibilities and associated probabilities accordingly.

The major difficulty with this approach is how to accurately calculate a points probability based on the available information. In particular, effective and efficient ways for applying spatial cohesion are required. In the work of Forne and Hayes [11] this is done by simply increasing a points likelihood based on the number of opaque neighbours it has. This works moderately well in some situations, but the obtained probabilities are often inaccurate, leading to reconstruction errors.

3 Scene Reconstruction

Like many tasks in image and vision computing, scene reconstruction is an ill-posed problem with inherent ambiguities. The basic objective of scene reconstruction is to recover or infer information about a scene from sensor data and any additional prior information. Because of sensor noise, modelling errors and a loss of information when projecting a three-dimensional scene into two-dimensions, an exact reconstruction is not possible. To deal with this a Bayesian approach can be used, where the objective is to find the most likely estimate of the scene given a set of measurements (in this case camera images) and any prior information.

Using $S = \{s_1, s_2, ..., s_L\}$ to represent the desired set of scene parameters and $I = \{I_1(x,y), I_2(x,y), ..., I_N(x,y)\}$

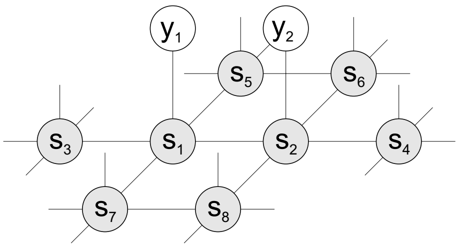

Figure 1: Pairwise Markov Random Field used to model the joint probability distribution of the system. Grey nodes are hidden scene variables. White nodes are the observation variables.

to represent the set of camera images, the reconstruction problem can be expressed as, given I find the most likely estimate of S. From Bayes rule, this can be expressed as

$$\max_{\text{arg S}} P(S|I) = \max_{\text{arg S}} \left(\frac{P(I|S)P(S)}{P(I)} \right). \quad (1)$$

The denominator, $P(I)$, being independent of our estimate $\hat{S}$ can be removed from the expression, giving

$$\max_{\text{arg S}} P(S|I) = \max_{\text{arg S}} \left(P(I|S)P(S) \right). \quad (2)$$

This states that the most likely scene is the one which maximises the joint probability distribution of the system.

In principle this problem is trivial and can be solved by searching through all possible combinations to find the one which is most likely. However, even for very small systems this approach is usually infeasible as the number of possible combinations is enormous.

3.1 System Model

In order to estimate and describe properties of a scene, such as structure and colour, and relate these to the observed sensor data, a model or representation of the physical system is required. In this paper, we have chosen to model the scene as a piece-wise continuous surface $z = z(x,y)$. This is represented using a 2-D array of variables $S = \{s_1, s_2, ..., s_L\}$, whos states correspond to possible surface heights. Although not a true volumetric representation, this model allows us to make a direct comparison between the work of Sun et al. [2] and our proposed greedy belief propagation algorithm. As with their work, the image data is modelled using an additional set of observation variables $Y = \{y_1, y_2, ..., y_L\}$ which represent the data associated with each scene variable.

The joint probability distribution of the system is then defined using a hidden pairwise Markov Random Field, as shown in Figure 1. Using this model, the joint probability function of this system is given by

$$P(S,Y) \propto \prod_i \psi_i(s_i, y_i) \prod_i \prod_{j \in N(i)} \psi_{ij}(s_i, s_j), \quad (3)$$

where $N(i)$ are the neighbours of i, $\psi_{ij}(s_i, s_j)$ is the compatibility function between neighbouring nodes s_i, s_j and $\psi_i(s_i, y_i)$ is the local evidence for node s_i, as determined from the data.

Relating this to the joint probability distribution $P(I,S) = P(I|S)P(S)$, the compatibility functions can be expressed as,

$$\prod_i \psi_i(s_i, y_i) \propto P(I|S) \quad (4)$$

and

$$\prod_i \prod_{j \in N(i)} \psi_{ij}(s_i, s_j) \propto P(S). \quad (5)$$

3.2 Conditional probability

Modelling the system noise as additional independent noise at each of the sensors, the conditional probability of obtaining I given S can be expressed as

$$P(I|S) = \prod_{j=1}^{N} \prod_{k=1}^{M} P(I_j(k)|I'_j(k)), \quad (6)$$

where $I_j(k)$ is the observed kth pixel intensity in the jth image and $I'_j(k)$ is the corresponding ideal image intensity, as would be obtain if the estimate was observed under noiseless conditions.

By modelling the noise distribution as a robust gaussian probability distribution, Eq (6) can be expressed as

$$P(I|S) = \prod_{j=1}^{N} \prod_{k=1}^{M} \left(\frac{1-\gamma}{\sigma\sqrt{2\pi}} \exp \frac{-\left[I'_j(k) - I_j(k) \right]^2}{2\sigma^2} + \gamma \right), \quad (7)$$

where σ is the variance of the gaussian distribution and γ is the robustness term.

If we now define a complete set of surface points $s_i(h_i) \in C$, as one whos forward projection completely fills image space, Eq (7) can be re-expressed as

$$P(I|S) = \prod_{s_i(h_i) \in C} V(s_i(h_i)), \quad (8)$$

where,

$$V(s_i(h_i)) = \prod_{j \in \Omega_{ih}} \left(\frac{1-\gamma}{\sigma\sqrt{2\pi}} \exp \frac{-\left[I'_j(\tau_{ihj}) - I_j(\tau_{ihj}) \right]^2}{2\sigma^2} + \gamma \right). \quad (9)$$

Here, Ω_{ih} is the subset of images which observe each point $s_i(h_i)$ and τ_{ihj} is the projected position of point $s_i(h_i)$ in image j. By taking the exponential term outside the function and replacing multiplication with summation, the probability of getting the image data corresponding to point $s_i(h_i)$ can be approximated by,

$$V(s_i(h_i)) = \exp\left(-\frac{1}{2\sigma^2}\sum_{j\in\Omega_{ih}}\left[I'_j(\tau_{ihj}) - I_i(\tau_{ihj})\right]^2\right) + e_u, \tag{10}$$

where e_u is the robust shaping parameter. Maximising this expression over all possible intensities, gives

$$I'_j(\tau_{ihj}) = \frac{1}{|\Omega_{ih}|}\sum_{j\in\Omega_{ih}} I_i(\tau_{ijh}) = \mu_{ih}, \tag{11}$$

where μ_{ih} is the mean observed intensity of surface point $s_i(h_i)$.

Finally substituting Eqs (8), (10) and (11) into Eq (4) and requiring that the scene is complete, the local compatibility functions are given by

$$\psi_i(s_i(h_i), y_i) = \exp\left(-\frac{1}{2\sigma^2}\sum_{j\in\Omega_{ih}}\left[\mu_{ih} - I_i(\tau_{ihj})\right]^2\right) + e_u, \tag{12}$$

3.3 Prior Knowledge

To improve the estimation process, prior knowledge about a scene can be incorporated into the reconstruction process. Although many possible priors can be used, we will only consider surface visibility and spatial cohesion. This simplifies the problem, making it easier to compare results between different algorithms.

The first of these, surface visibility, is included so as to favour more visible surfaces. For implementational reasons this term is actually incorporated into the local compatibility by replacing the term,

$$\frac{1}{2\sigma^2}\sum_{j\in\Omega_{ih}}\left[\mu_{ih} - I_i(\tau_{ihj})\right]^2, \tag{13}$$

with,

$$\frac{1}{|\Omega_{ih}|}\left(\frac{1}{2\sigma^2}\sum_{j\in\Omega_{ih}}\left[\mu_{ih} - I_i(\tau_{ijh})\right]^2\right) + \frac{(N - |\Omega_{ih}|)}{e_o}, \tag{14}$$

where $(N - |\Omega_{ih}|)$ is the number of images in which point $s_i(h_i)$ is occluded and e_o is related to the probability of occlusion.

The second term, spatial cohesion, is based on the fact that scenes are more likely to consist of a few continuous surfaces, rather than a random cloud of points. This is incorporated into the Markov Random Field by defining the compatibility functions between neighbouring nodes as

$$\psi_{ij}(s_i, s_j) = \max\left(\exp\left(\frac{-|s_i - s_j|}{\sigma_d}\right), e_d\right), \tag{15}$$

where σ_d and e_d are shaping parameters describing the probability distribution.

4 Greedy Belief Propagation

Having modelled the system as a pairwise MRF, the optimal scene estimate is obtained by maximising the corresponding joint probability distribution. To do this we propose a new greedy belief propagation algorithm. This is basically the same as standard Belief Propagation except the local compatibility functions are updated to reflect changes in the conditional probability distribution. These updates are performed in a similar manner to that used by the Greedy algorithm of Forne and Hayes [11]. Beginning with an initially transparent estimate, surface points are progressively assigned as opaque until a complete scene estimate has been formed. However, unlike the greedy algorithm, the assignment of points is reversible, so that points may be restored to a transparent state.

To begin with, the local compatibility functions $\psi_i(s_i, y_i)$ are calculated assuming all surfaces are visible to all cameras. Standard Belief Propagation is then performed on the belief network for a fixed number of iterations. Next, the most likely set of surface points are assigned to be opaque. This is done by simply selecting those points whose beliefs are above some threshold value. Scene visibilities are then updated along with the associated set of local compatibility functions. Following this, Belief Propagation is continued for another fixed number of iterations. This process is repeated until a complete estimate of the scene is formed.

During each update, the algorithm checks all currently assigned points to see if their belief has fallen below the threshold. If this is the case, the corresponding scene point is restored to a transparent state, thereby allowing decisions to be undone.

To ensure convergence, the threshold is progressively lowered until all nodes have been assigned some height.

5 Experimental Results

To test the performance of the proposed greedy belief propagation algorithm, two small synthetic test sets were generated. These were chosen to demonstrate the problems with existing algorithms and to show how the proposed algorithm can help to overcome them.

The first test scene consisted of two square planes located at different heights above a third background plane. A sequence of four ideal images of the scene were then generated based on a square arrangement of cameras. To each image gaussian noise was added, giving a 25dB signal to noise ratio, simulating camera noise and distortions. The resulting top lefthand image from this sequence is shown in Figure 2(a). This was also used as the reference image for the various reconstruction algorithms. The ideal depth map for this scene is shown in Figure 2(b).

The second test scene comprised of three circular planes located at different heights above a fourth background plane. As with the first test set, a sequence of four ideal images of the scene were generated based on a square arrangement of cameras. Gaussian noise was added to each image, again giving a 25dB signal to noise ratio. The resulting top lefthand image from this test set is shown in Figure 3(a), while the ideal depth map for this image is shown in Figure 3(b).

The Greedy Belief Propagation algorithm was then tested using the same set of parameters for both test cases. The gaussian noise variance σ was set to 10, the uniform noise parameter e_u to 0.02, the visibility parameter e_o to 0.1, the smoothing parameter σ_d to 0.7 and the smoothing parameter e_d to 0.05. The number of belief propagation iterations between scene updates was also fixed to 4.

Results from the Greedy Belief Propagation algorithm are shown in Figures 2(c) and 3(c).

To observe the effect updating the local compatibility functions has on the results, a standard Belief Propagation algorithm was tested on the data assuming full visibility. Results from this are shown in Figures 2(d) and 3(d). An implementation of Sun et al's. algorithm [2] suitable for a 2-D array of cameras was also tested, with results shown in Figures 2(e) and 3(e). Finally the results were compared to the greedy algorithm proposed by Forne and Hayes [11]. These are shown in Figures 2(f) and 3(f).

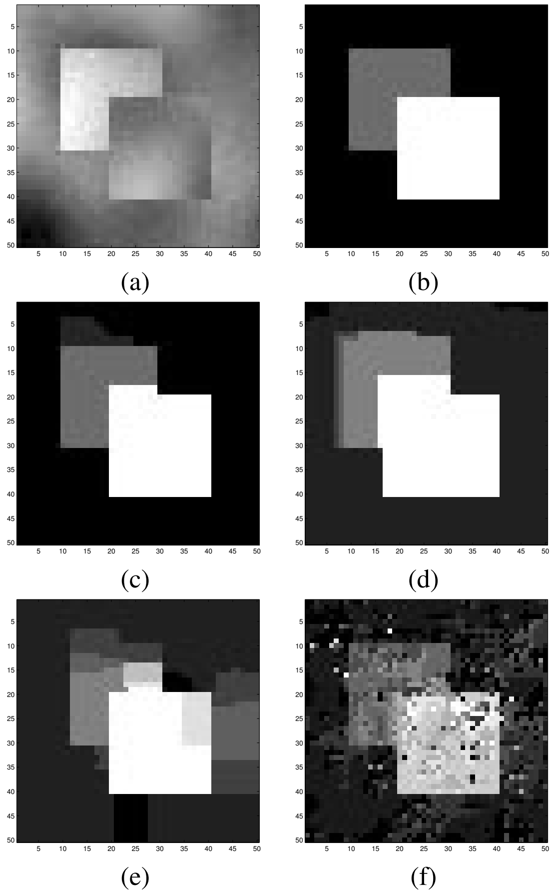

Figure 2: Test set 1. (a) Top lefthand image from sequence. (b) Ideal depth map. (c) Results from Greedy belief propagation. (d) Results from standard BP algorithm, assuming full visibilities. (e) Results from implementation of Sun's et al's. BP algorithm. (f) Results from greedy algorithm.

6 Discussion

Comparing the results from Greedy Belief Propagation with those obtained from standard BP and Sun et al's. algorithm it is apparent how important accurate compatibility functions are. As shown in both of the test examples, inaccurate compatibility functions can lead to errors which extend well beyond the occluded regions. The proposed Greedy Belief Propagation algorithm does a good job of ensuring that these are consistent with the estimated scene, leading to improved results on both test sets.

One of the main problems experienced with the Greedy Belief Propagation algorithm was the assignment of incorrect points. By assigning a different set of opaque points to those present in the actual scene, an incorrect set of local compatibility functions will be formed. This can then cause the Belief Propagation process to be lead astray. Usually this was not too much of a problem as points incorrectly assigned at one iteration were typically reversed on the subsequent iteration.

To address this problem a more reliable measure of probability is required. Unfortunately one of the problems with Belief Propagation is that the calculated beliefs for each node give a measure of how each node effects the joint probability distribution of the whole system rather than how likely or otherwise the point is.

7 Conclusions and Future Work

One of the fundamental problems with scene reconstruction is that the posteriori probability of a scene point depends not only on the image data mapping to that point but also on it visibility. Unfortunately these are unknown prior to reconstruction and depend on the rest of the scene. Because of this dependence, it is

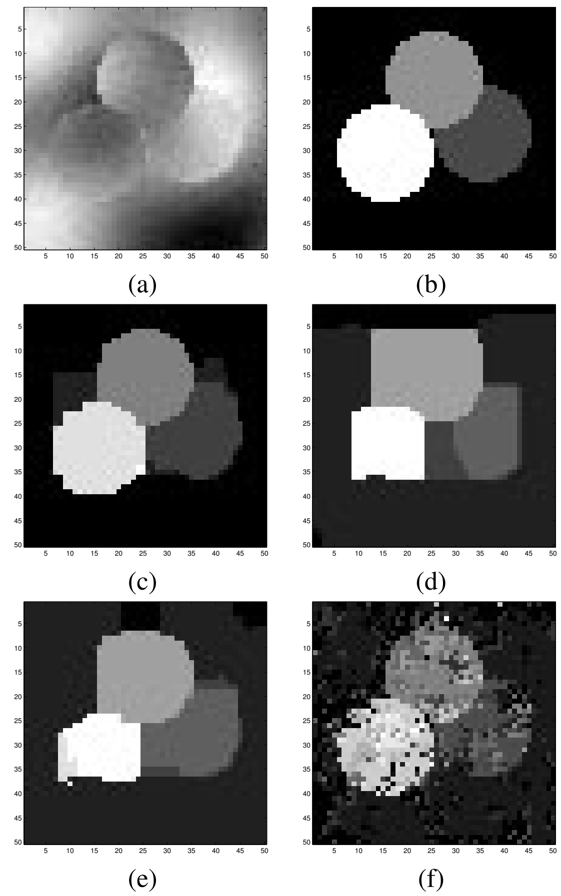

Figure 3: Test set 2. (a) Top lefthand image from sequence. (b) Ideal depth map. (c) Results from Greedy belief propagation. (d) Results from standard BP algorithm, assuming full visibilities. (e) Results from implementation of Sun's et al's. BP algorithm. (f) Results from greedy algorithm.

difficult to produce an accurate measure for the likelihood of a point unless the scene is considered as a whole. To address this difficulty a novel Greedy Belief Propagation algorithm is presented which progressively updates the local probability functions associated with each point as the scene estimate is refined. This offers many of the advantages of standard Belief Propagation, while at the same time dealing with the visibility problem. Results from synthetic test sets demonstrate the advantages of this approach, and highlight the problems caused by unknown visibilities.

Future work will focus on improving the system model as well as attempting to speed up the existing algorithm. It will also investigate methods for making the assignment of opaque points more reliable.

8 Acknowledgements

This work was supported by a Top Achiever Doctoral Scholarship from the Foundation of Research, Science and Technology, New Zealand.

References

[1] R. Lane and N. Thacker. Stereo vision research: An algorithm survey, 1996.

[2] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800, July 2003.

[3] Vladimir Kolmogorov, Ramin Zabih, and Steven Gortler. Generalized multi-camera scene reconstruction using graph cuts. In *Proceedings of Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume LNCS, 2683, pages 501–516, July 2003.

[4] Georgy Gimel'farb. Stereo terrain reconstruction by dynamic programming. Technical Report CITR-TR-21, Center for Image Technology and Robotics, Computer Science Department, University of Auckland, 1998.

[5] B. Culbertson, T. Malzbender, and G. Slabaugh. Generalized voxel coloring. In *Proceedings of the ICCV'99 Vision Algorithms Workshop*, September 1999.

[6] C. Harding and R. Lane. Passive navigation from image sequences by use of a volumetric approach. *Journal of the Optical Society of America A*, 19(2):295–305, February 2002.

[7] Georgy Gimel'farb. Stereo terrain reconstruction by dynamic programming, 1997.

[8] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7–42, April-June 2002.

[9] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, California, revised second edition, 1988.

[10] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer. A survey of volumetric scene reconstruction methods from photographs. Technical Report 1, Center for Signal and Image Processing, Georgia Institute of Technology, February 2001.

[11] C. Forne and M. Hayes. A maximum likelihood approach to reconstructing scenes from photometric images. In *Proceedings of Image and Vision Computing New Zealand Conference*, pages 151–156, November 2001.