An MRF Optimisation Framework for Full 3D Helmholtz Stereopsis

Gianmarco Addari and Jean-Yves Guillemaut
Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, U.K.

Keywords: 3D Reconstruction, Helmholtz Stereopsis, Markov Random Fields.

Abstract: Accurate 3D modelling of real world objects is essential in many applications such as digital film production and cultural heritage preservation. However, current modelling techniques rely on assumptions to constrain the problem, effectively limiting the categories of scenes that can be reconstructed. A common assumption is that the scene's surface reflectance is Lambertian or known a priori. These constraints rarely hold true in practice and result in inaccurate reconstructions. Helmholtz Stereopsis (HS) addresses this limitation by introducing a reflectance agnostic modelling constraint, but prior work in this area has been predominantly limited to 2.5D reconstruction, providing only a partial model of the scene. In contrast, this paper introduces the first Markov Random Field (MRF) optimisation framework for full 3D HS. First, an initial reconstruction is obtained by performing 2.5D MRF optimisation with visibility constraints from multiple viewpoints and fusing the different outputs. Then, a refined 3D model is obtained through volumetric MRF optimisation using a tailored Iterative Conditional Modes (ICM) algorithm. The proposed approach is evaluated with both synthetic and real data. Results show that the proposed full 3D optimisation significantly increases both geometric and normal accuracy, being able to achieve sub-millimetre precision. Furthermore, the approach is shown to be robust to occlusions and noise.

1 INTRODUCTION

Many industries such as film, gaming and cultural heritage require the ability to accurately digitise real world objects. Despite significant progress in 3D modelling over the past decades, modelling scenes with complex surface reflectance (e.g. glossy materials or materials with spatially varying and anisotropic reflectance properties) remains an open problem. Existing modelling techniques usually make simplifying assumptions on the scene's Bidirectional Reflectance Distribution Function (BRDF) which fail to capture the complexity of natural scenes. For instance, multi-view stereo reconstruction techniques rely on the assumption that the scene is Lambertian or sufficiently textured to be able to use photo-consistency to infer image correspondences across views. Photometric Stereo (PS) can handle more complex types of surface reflectance, but requires prior knowledge of the BRDF, which is very difficult to acquire in practice.

A solution to generalise 3D modelling to scenes with complex reflectance has been proposed in the form of Helmholtz Stereopsis (HS). The approach utilises the principle of Helmholtz reciprocity to derive a reconstruction methodology which is independent of the scene's surface reflectance. While very promising results were demonstrated, previous formulations of HS have been mostly limited to 2.5D reconstruction (Zickler et al., 2002) (Roubtsova and Guillemaut, 2018), thereby providing only a partial scene reconstruction. Recently in (Delaunoy et al., 2010), an approach was proposed to extend HS to the 3D domain, however, being based on gradient descent, the approach is dependent on a good initialisation to ensure convergence to the global optimum. Overall, full 3D scene reconstruction using HS has received limited consideration, with MRF formulations limited to 2.5D scenarios.

This paper advances the state-of-the-art in modelling of scenes with arbitrary unknown surface reflectance by proposing the first Markov Random Field (MRF) formulation of HS for full 3D scene reconstruction. The paper makes two key contributions in this area. First, it introduces a novel pipeline for full 3D modelling via fusion of 2.5D surface reconstructions obtained from multiple view-points through MRF optimisation. Second, it proposes a volumetric MRF formulation of HS that permits direct optimisation of the complete 3D model and can be used to refine the previous estimate. Being based on an MRF formulation, the approach is able to cope with coarse initialisation and is robust to noise.

Addari, G. and Guillemaut, J.
An MRF Optimisation Framework for Full 3D Helmholtz Stereopsis.
DOI: 10.5220/0007407307250736
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 725-736
ISBN: 978-989-758-354-4

2 RELATED WORK

The most common approaches to 3D reconstruction are Shape from Silhouettes (SfS), multi-view stereo and PS.

SfS was proposed in (Laurentini, 1994) and consists in using 2D silhouette data to reconstruct a Visual Hull (VH) of the object. Despite having recently been improved upon in recent works (Liang and Wong, 2010) (Nasrin and Jabbar, 2015), SfS suffers from being unable to reconstruct concavities.

Classic binocular and multi-view stereo approaches (Szeliski et al., 2008) (Seitz et al., 2006) allow for more complex geometric reconstructions than SfS, but are limited by the scene BRDF, which needs to be Lambertian. As this is often not the case, assuming the wrong BRDF can lead to incorrect reconstructions. Some more recent multi-view techniques (Nishino, 2009) (Oxholm and Nishino, 2016) (Lombardi and Nishino, 2016) attempt to jointly estimate the geometry and reflectance models of the scene by calculating iteratively the scene shape from the current estimated reflectance and vice versa, which ultimately constrains each calculation to the accuracy with which the other parameter was estimated.

Finally, PS (Woodham, 1980) consists in computing the normals of a scene given a set of inputs with stationary point of view but varying lighting. Despite the possibility of reconstructing non-Lambertian scenes, in PS the BRDF needs to be known a priori. State-of-the-art work includes (Vogiatzis et al., 2006), where a shiny, textureless object is reconstructed using shadows and varying illumination, under the simplifying assumption of a Lambertian reflectance model. In (Chandraker et al., 2013) image derivatives are used to reconstruct surfaces with unknown BRDF, albeit being limited to isotropic BRDFs. In (Goldman et al., 2010) a generic form of BRDF is used to compute shapes, each point on the surface is considered to be a mixture of previously calculated BRDFs, however this is limited to a maximum of two materials per point. In (Han and Shen, 2015), PS is computed for isotropic and anisotropic reflectance models by considering the following characteristics of the BRDF: the diffuse component, the concentration of specularities and the resulting shadows. Despite achieving great results, the set-up in this paper is extremely complex and a large number of different light directions are needed to perform the surface reconstruction.

The use of Helmholtz reciprocity (Von Helmholtz and Southall, 1924) for 3D reconstruction was first proposed by Magda et al. in (Magda et al., 2001) where the principle is used to recover the geometry of scenes with arbitrary and anisotropic BRDF in a simplified scenario. The proposed technique was then further developed into HS in (Zickler et al., 2002) by utilising it to perform normal estimation as well. In the classical HS formulation maximum likelihood is used to determine the depth of each point in the scene, while the normals are estimated using Singular Value Decomposition (SVD). No integration is performed over the surface, resulting in discontinuities and a noisy reconstruction. To enforce Helmholtz reciprocity, a set-up similar to the one shown in Figure 1 needs to be used. Let us consider a camera and an isotropic, unit-strength point light source, respectively positioned at $\boldsymbol{O}_l$ and $\boldsymbol{O}_r$. For a given surface point $\boldsymbol{P}$, the irradiance measured at its projection in the left image (i_l) can be expressed as:

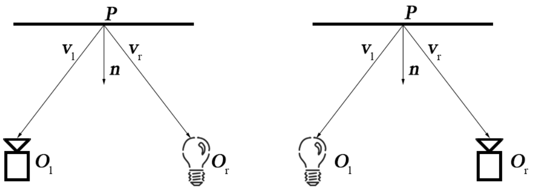


Figure 1: Camera/light pair positioning in HS.

$$i_l = f_r(\boldsymbol{v}_r, \boldsymbol{v}_l) \cdot \frac{\boldsymbol{n} \cdot \boldsymbol{v}_r}{|\boldsymbol{O}_r - \boldsymbol{P}|^2} \qquad (1)$$

where f_r indicates the BRDF at point $\boldsymbol{P}$ with incident light direction $\boldsymbol{v}_r$ and viewing direction $\boldsymbol{v}_l$, $\boldsymbol{n}$ is the surface normal at $\boldsymbol{P}$ and $\frac{1}{|\boldsymbol{O}_r - \boldsymbol{P}|^2}$ accounts for the light falloff. If the light and camera are interchanged, a corresponding equation for the irradiance of the projection of $\boldsymbol{P}$ in the right image (i_r) is obtained where $f_r(\boldsymbol{v}_l, \boldsymbol{v}_r) = f_l(\boldsymbol{v}_r, \boldsymbol{v}_l)$, due to the Helmholtz reciprocity constraint. By performing a substitution, the reflectance term can be removed and the following equation, independent from the surface BRDF, is obtained:

$$\left(i_l \frac{\boldsymbol{v}_l}{|\boldsymbol{O}_l - \boldsymbol{P}|^2} - i_r \frac{\boldsymbol{v}_r}{|\boldsymbol{O}_r - \boldsymbol{P}|^2} \right) \cdot \boldsymbol{n} = \boldsymbol{w} \cdot \boldsymbol{n} = 0 \qquad (2)$$

Utilising multiple camera light pairs (at least three) allows to obtain a matrix W where each row is a $\boldsymbol{w}$ vector. By minimising the product $W \cdot \boldsymbol{n}$ it is possible to obtain an estimate of the normal at point $\boldsymbol{P}$. To do so the matrix is decomposed using SVD:

$$SVD(W) = U \Sigma V^T \qquad (3)$$

where Σ is a diagonal matrix and U and V are orthogonal matrices. The last column of V gives an estimate of the normal, while the non zero terms in the diagonal matrix $(\sigma_1, \sigma_2, \sigma_3)$ can be used to compute a quality measure of the normal.

In (Zickler et al., 2003) and (Tu and Mendonca, 2003) it is demonstrated that HS can be performed

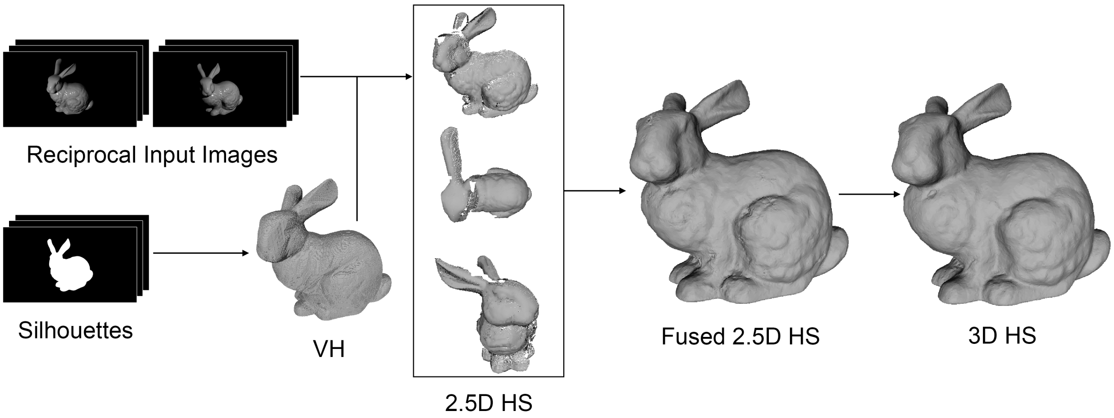


Figure 2: Pipeline overview.

with as low as a single pair of reciprocal images, however the assumption of a C_1 continuous surface is made, making it impossible to reconstruct surfaces that present discontinuities. Further work include (Guillemaut et al., 2004) where HS is applied to rough and textured surfaces by integrating the image intensities over small areas, (Janko et al., 2004) where it is shown how performing radiometric scene calibration allows for a vast improvement in the normal accuracy calculated using HS, (Zickler, 2006) where geometric and radiometric calibration are performed by exploiting correspondences between the specular highlights in the images and finally (Guillemaut et al., 2008) where an alternative radiometric distance is proposed to perform a maximum likelihood surface normal estimation. All these methods are performed in 2.5D and do not attempt to compute a globally optimal surface, calculating instead an occupancy likelihood point by point.

In (Roubtsova and Guillemaut, 2017) coloured lights are used to reconstruct dynamic scenes using only three cameras and a Bayesian formulation is proposed to perform a globally optimal reconstruction of the scene, while in (Roubtsova and Guillemaut, 2018) the maximum a posteriori formulation is applied to classic HS by enforcing consistency between the points depth and the estimated normals. This allows to obtain less noisy results, however the scope of these works is still restricted to 2.5D surfaces and has not been applied to full 3D scenes. Furthermore, occlusions are not handled by this method, affecting its performance and severely restricting the scope of scenes that can be reconstructed.

HS was first applied in the 3D domain in (Weinmann et al., 2012), where it is used to complement structured light consistency, which is unable to obtain high frequency details on the reconstructed surface and (Delaunoy et al., 2010), where a variational formulation is presented to reconstruct a 3D mesh using gradient descent optimisation. In Weinmann's work HS is only used as a refinement step on areas of the surface where fine details are present and a very complex set-up consisting of a light dome is used, which makes this method extremely difficult to deploy and constrained to a specific set of scenes. In Delaunoy's work, instead, the set-up is simply a turntable, a pair of fixed lights and a fixed camera, which makes it easier to reproduce. However, because the optimisation employed is based on gradient descent, global optimality is not guaranteed and the method could get trapped in local minima if a proper initialisation is not provided.

In contrast, the proposed method performs multiple 2.5D reconstructions from different view-points, which are then fused together to obtain a full 3D model of the scene. This is followed by an MRF optimisation on the 3D reconstruction, which, compared to gradient descent, is less reliant on having a good initialisation and benefits from some optimality guarantees depending on the choice of algorithm used to optimise the energy function. During both steps self-occlusions are taken into consideration by performing an approximate visibility check.

3 PIPELINE OVERVIEW

In this section the pipeline used to perform the full 3D reconstruction is broken down, and each step is detailed. As shown in Figure 2, the inputs to this method are calibrated Helmholtz reciprocal pairs of images of the object and the silhouettes for each view. The camera positions are arbitrary. The first step consists in defining a voxel grid that contains the whole object, and reconstructing its VH applying SfS to the silhouettes.

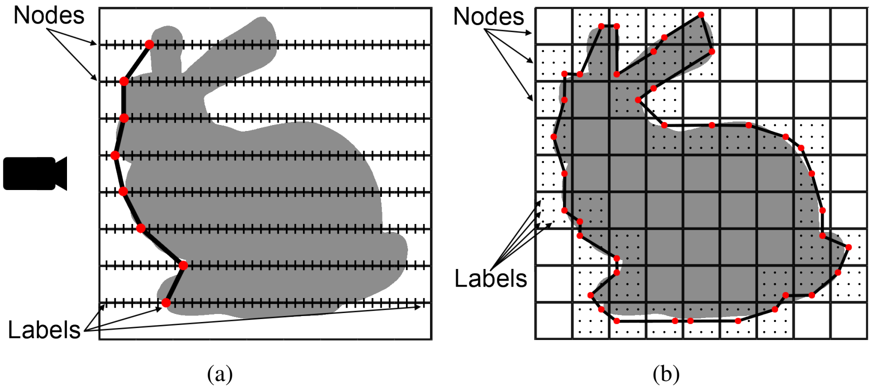


Figure 3: Simplified representation of labelling in the 2.5D (a) and 3D (b) methods.

The VH is then used to initialise the next step, where a set of separate views of the object are reconstructed using a visibility aware Bayesian formulation of HS. The approach extends the method from (Roubtsova and Guillemaut, 2018) with the use of additional information on visibility provided by the VH to select a subset of cameras for each view and point independently. Selecting camera visibility correctly is a critical step since the cameras are not all placed on a plane as in other 2.5D Helmholtz methods, but can instead be placed anywhere in the 3D space surrounding the object, depending on how the dataset was collected. In the proposed approach, reconstruction is performed from the six viewing directions defined by the cardinal axes of the reference frame. These define six orthographic virtual cameras and provide sufficient coverage to reconstruct the complete scene.

To remove redundancies and possible inconsistencies among the partial surfaces, they are fused together using Poisson surface reconstruction. The resulting surface is then used to initialise the final step of the reconstruction pipeline. In this final step, the problem is defined as a volumetric MRF, where each voxel corresponds to a node and the labelling defines whether the node is outside, inside, or on the surface of the reconstructed scene. A refined 3D model is obtained by MRF optimisation using a tailored Iterative Conditional Modes (ICM) algorithm.

4 2.5D RECONSTRUCTION FUSION

During this step multiple 2.5D depth maps of the object are computed from different directions using HS and optimised using MRF. Each problem is formulated on an orthographic grid where each node corresponds to an image point in the reference frame of a virtual camera aligned with the chosen direction. Each node can be assigned a label which indicates the depth at which the surface is located at the corresponding pixel as shown in Figure 3a. The set of labels is $l_0,...,l_{d-1}$, where l_0 indicates the point in the grid closest to the virtual camera and l_{d-1} indicates the farthest point in the grid. Each node is assigned a label to minimise the following energy function:

$$E(l) = (1-\alpha)\sum_{p\in I} D_{2D}(B(p,l_p)) + \alpha \sum_{p,q\in\mathcal{N}_{2D}} S_{2D}(B(p,l_p),B(q,l_q)) \quad (4)$$

where α is a balancing parameter between the data and smoothness terms, I is the 2-dimensional grid defined by the virtual camera, $D_{2D}(B(p,l_p))$ is the data term of the function, which corresponds to the Helmholtz saliency measured at 3D point $B(p,l_p)$, obtained by back-projecting image point $\boldsymbol{P}$ at the depth corresponding to label l_p. $\mathcal{N}_{2D}$ indicates the neighbourhood of a node, which consists of the four pixels directly adjacent in the image and $S_{2D}(B(p,l_p),B(q,l_q))$ is the smoothness term, and it corresponds to the normal consistency term calculated between 3D points $B(p,l_p)$ and $B(q,l_q)$.

In this formulation the data term is computed using the following equation:

$$D_{2D}(\boldsymbol{P}) = \begin{cases} 1, & \text{if } |vis(\boldsymbol{P})| < min_{vis} \\ e^{-\mu\times\frac{\sigma_2(\boldsymbol{P})}{\sigma_3(\boldsymbol{P})}}, & \text{otherwise} \end{cases} \quad (5)$$

where $vis(\boldsymbol{P})$ indicates the set of reciprocal pairs of cameras from which point $\boldsymbol{P}$ is visible, min_{vis} is a variable set to enforce the minimum number of reciprocal pairs of cameras that make a normal estimate reliable, μ is assigned the value $0.2\ln(2)$ to replicate the same weight used in (Roubtsova and Guillemaut, 2018) and σ_2 and σ_3 are two values from the diagonal matrix obtained performing SVD on matrix W as shown in Equation 3.

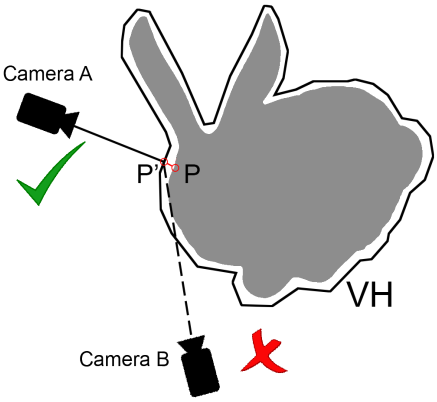


Figure 4: Points on the surface (***P***) are approximated to the closest point on the VH (***P'***) before occlusions are taken into consideration for visibility.

An important contribution to enable application to complex 3D scenes is the introduction of the visibility term in the formulation. The first criterion to determine visibility is to only consider the cameras whose axis stands at an angle smaller than 80° with respect to the virtual camera axis. Then occlusions are computed by approximating each point's visibility based on the visibility of its closest point on the surface of the VH. If an intersection is found between the VH and the segment connecting the camera center to the approximated point, the camera and its reciprocal are considered to be occluded and therefore are not used as shown in Figure 4.

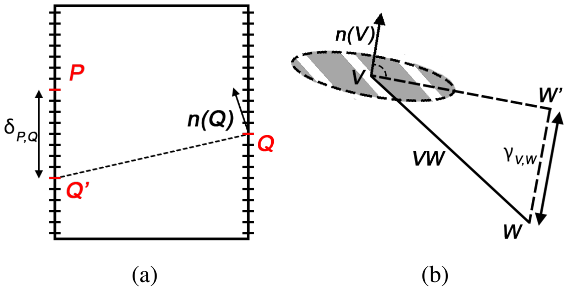
The smoothness function used here is the distance based DNprior (Roubtsova and Guillemaut, 2018), which enforces a smooth surface that is consistent with the normals obtained through HS. This term is calculated as follows:

$$S_{2D}(\boldsymbol{P},\boldsymbol{Q}) = \begin{cases} \frac{1}{2}(\delta^2_{\boldsymbol{P},\boldsymbol{Q}}+\delta^2_{\boldsymbol{Q},\boldsymbol{P}}), & \text{if } \delta_{\boldsymbol{P},\boldsymbol{Q}} \text{ and } \delta_{\boldsymbol{Q},\boldsymbol{P}} < t \\ t^2, & \text{otherwise} \end{cases} \tag{6}$$

where t is the maximum threshold for $\delta_{\boldsymbol{P},\boldsymbol{Q}}$. $\delta_{\boldsymbol{P},\boldsymbol{Q}}$ is the distance between point ***P*** and the projection of ***Q***, perpendicular to its estimated normal, on the grid pixel where ***P*** lies as illustrated in Figure 5a, and it is calculated as follows:

$$\delta_{\boldsymbol{P},\boldsymbol{Q}} = \frac{\boldsymbol{PQ}\cdot\boldsymbol{n(Q)}}{\boldsymbol{n(Q)}\cdot\boldsymbol{C}} \tag{7}$$

where ***PQ*** is the vector connecting ***P*** and ***Q***, ***n(Q)*** indicates the estimated normal at point ***Q*** and ***C*** is the virtual camera axis. Whenever $\delta_{\boldsymbol{P},\boldsymbol{Q}}$ or $\delta_{\boldsymbol{Q},\boldsymbol{P}}$ are greater than a threshold t dependent on the reconstruction resolution, this term is truncated to t^2 in order to avoid heavy penalties where a strong discontinuity is present on the surface.



(a) (b)

Figure 5: Illustration of how S_{2D} (a) and S_{3D} (b) are computed.

The energy function is then minimised using Tree Reweighted Message Passing (TRW) (Kolmogorov, 2006) to obtain the depth maps from each viewing direction. These are then fused together using Poisson surface reconstruction (Kazhdan et al., 2006), and the result is used to initialise the next step in the pipeline.

5 FULL 3D OPTIMISATION

In this step the reconstruction is further optimised in full 3D. This is done by enforcing coherence between neighbouring points in 3D, which is not guaranteed when fusing partial depth maps obtained from different points of view. A 3D orthographic grid is created that encompasses the whole object, and each voxel is assigned a node in a multi-labelled MRF graph. The labels are the following: $\{I, O, L_0, \dots, L_{R-1}\}$, where I and O indicate respectively whether the voxel is inside or outside the reconstructed surface, while the remaining labels are assigned when the node is on the surface. More specifically, the surface labels $\{L_0, \dots, L_{R-1}\}$ characterise the position of the surface point in the local reference frame of each voxel. In this paper, surface labels are defined by regularly sampling the interior of a voxel as shown in 3b, although other sampling strategies could also be considered. In this MRF formulation the following energy function is minimised to optimise the surface:

$$\begin{aligned} E(L) = (1-\beta)\sum_{v\in\mathcal{V}} D_{3D}(v, L_v) + \\ \beta \sum_{v,w\in\mathcal{N}_{3D}} S_{3D}(v, L_v, w, L_w) \end{aligned} \tag{8}$$

where β is a weight to balance the effects of the data and smoothness terms, $\mathcal{V}$ is the 3D grid and $\mathcal{N}_{3D}$ is the neighbourhood composed by the six voxels directly adjacent to the current one. $M(v, L_v)$ indicates the position of the surface point at node v when assigned label L_v. The data term is computed differently depending on whether the point is inside, outside or on the surface:

$$D_{3D}(v, L_v) \;=\; \begin{cases} 0, & L_v \in \{I, O\} \\ e^{-\mu \times \frac{\sigma_2(M(v,L_v))}{\sigma_3(M(v,L_v))}}, & \text{otherwise} \end{cases} \tag{9}$$

Similarly to the data term, the way the smoothness term is computed depends on the label combination of the two nodes:

$$S_{3D}(v, L_v, w, L_w) =$$
$$\begin{cases} \Gamma(M(v, L_v), M(w, L_w)), & L_v, L_w \in \{L_0, \ldots, L_{R-1}\} \\ \infty, & L_v, L_w \in \{I, O\},\ L_v \neq L_w \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where

$$\Gamma(\boldsymbol{V}, \boldsymbol{W}) = \frac{1}{2}(\gamma_{\boldsymbol{V},\boldsymbol{W}}^2 + \gamma_{\boldsymbol{W},\boldsymbol{V}}^2) \tag{11}$$

indicates the normal consistency in the 3D optimisation between points ***V*** and ***W***. $\gamma_{\boldsymbol{V},\boldsymbol{W}}$ is calculated as follows:

$$\gamma_{\boldsymbol{V},\boldsymbol{W}} = |\boldsymbol{VW} \cdot \boldsymbol{n(W)}| \tag{12}$$

where ***VW*** is the vector connecting points ***V*** and ***W***, while ***n(W)*** indicates the unit normal estimated via HS at point ***W***. This term consists of the distance between ***W*** and the plane perpendicular to ***n(V)*** intersecting point ***V***. In Figure 5b an illustration of how this term is calculated is shown. The ∞ term is here used to constrain inside and outside voxels to be separated by surface voxels, thus avoiding an empty solution where all nodes are either labelled to be inside or outside. The normal consistency term does not require truncation in the 3D domain.

Once the graph is initialised, the optimisation is performed using a tailored version of ICM (Besag, 1986). ICM is an exhaustive search algorithm that iterates through an MRF graph and changes one variable at a time by trying to optimise its local neighbourhood cost. In its classic formulation, ICM would not work in this scenario because of the constraint on the surface. Namely, changing the label of a surface node to be either outside or inside would result in a hole on the surface, which is currently prevented by having an infinite weight when outside and inside voxels are neighbours. However, by changing two neighbouring variables at a time, and considering all neighbouring nodes to at least one of the two variables, the surface can be shifted close to its optimal solution through multiple iterations. Only tuples where one node is on the current surface of the reconstruction are considered, and for each tuple and their neighbours, all possible configurations are considered, selecting the solution with the lowest energy. Since the problem is initialised close to the actual surface, this step typically converges after a small number of iterations. Finally, the nodes labelled to be on the surface are extracted together with their Helmholtz estimated normals and are integrated using Poisson surface reconstruction to obtain a mesh representation.

6 EVALUATION

In this section a brief description of the datasets used to perform the evaluation is given, followed by an analysis on the results obtained. The methods used will hereby be denoted as follows: 'VH' for the reconstruction obtained using SfS, '2.5D HS' for the 2.5D reconstructions obtained using the approach described in Section 4, 'Fused 2.5D HS' for the fusion of the 2.5D surfaces obtained using '2.5D HS' and '3D HS' for the proposed method that performs MRF optimisation on the full 3D surface. Where available, the ground truth is labelled 'GT'.

6.1 Dataset

To test the methodology, both synthetic and real scenes were used. It is important to note that, when generating synthetic scenes, commonly used rendering pipelines often break Helmholtz reciprocity, providing images that are not physically plausible. To address this, synthetic images were rendered using the modified Phong reflectance model (Lewis, 1994), which combines a diffuse and a specular part. In Figure 6 two reciprocal pairs of images are shown for the Stanford Bunny (Turk and Levoy, 1994) and Armadillo scenes (Krishnamurthy and Levoy, 1996). These scenes were chosen because they both present elongated thin structures, namely the ears of the bunny and the limbs and claws of the armadillo; strong specularities with no textures; and numerous self occlusions. The synthetic scenes were also distorted to measure the robustness of the methodology against noisy input data. In Figure 7 a close up of the distorted images used for the experiments are shown.

Each scene is composed of 40 reciprocal pairs of images captured from a set of viewpoints obtained sampling a sphere around the object. The images are rendered at a resolution of 1920 × 1080. Using synthetic scenes allows for a quantitative evaluation of the methodology, by comparing the results obtained against the ground truth data.

The real dataset from (Delaunoy et al., 2010) is composed of two scenes called Dragon and Fish. Two reciprocal pairs of images are shown in Figure 8. These two datasets are challenging due to the strong specularities present on the surface of 'Fish' and the

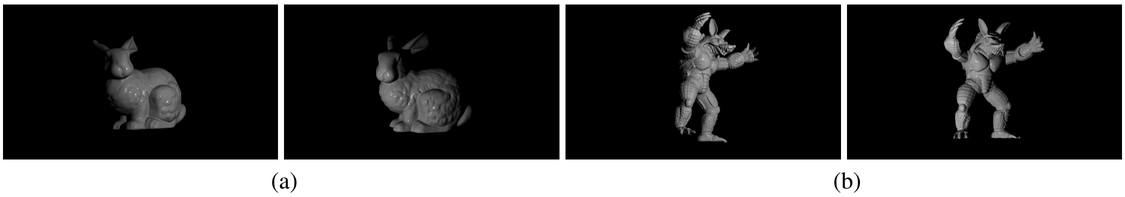

(a) (b)

Figure 6: Two reciprocal pairs from the synthetic dataset: ‘Bunny’ (a) and ‘Armadillo’ (b).

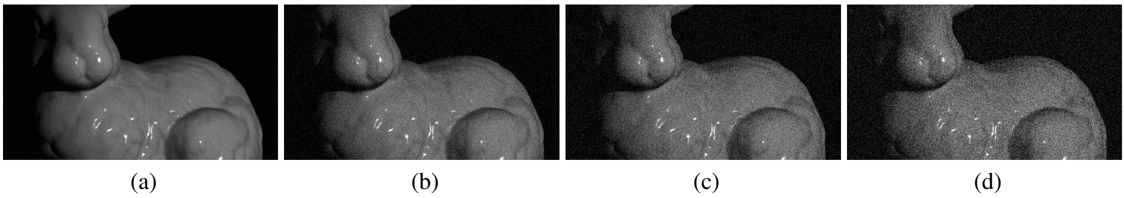

(a) (b) (c) (d)

Figure 7: Images with added input noise (white noise with normalised standard deviation 0.001 (a), 0.005 (b), 0.010 (c), 0.020 (d)).

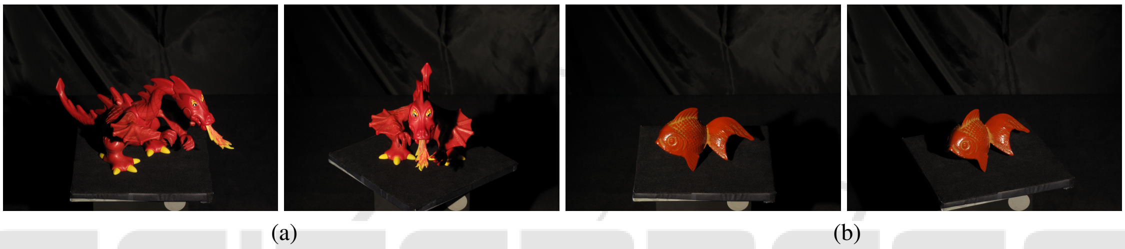

(a) (b)

Figure 8: Two reciprocal pairs from the real dataset: ‘Dragon’ (a) and ‘Fish’ (b).

numerous self occlusions in ‘Dragon’. It must be noted that no ground truth data or laser scan is available for these datasets, and thus only a qualitative evaluation could be performed. The resolution of the images from these scenes is 1104×828 and they are all from a single ring of cameras positioned on top of the objects, which means that one side of the object is always completely occluded. This is relevant to show how the reconstruction method used is able to deal with a lack of data.

6.2 Results

6.2.1 Synthetic Scenes

In this section a comparison between ‘VH’, ’Fused 2.5D HS’ and ‘3D HS’ is performed on the synthetic scenes to quantitatively assess the performance of the proposed approach compared to the state-of-the-art 2.5D technique. It was decided to compare the proposed methodology directly against ‘Fused 2.5D HS’ because ‘2.5D HS’ produces only partial reconstructions of the object using a limited number of views. The parameters used to measure performance are accuracy at 90% and completeness calculated at different thresholds depending on the scene examined. The thresholds are chosen depending on the VH accuracy, to measure whether there are holes in the reconstructed mesh or areas where the accuracy is significantly degraded.

In Figure 9 the results obtained from each method are shown. The results reported in this paper were obtained using the following parameters: $\{\alpha = 0.3, \beta = 0.4, t = 3 \times r\}$ where r is the edge length of a pixel in the reference frame. Starting from the left the initial VH is shown, followed by some of the results obtained using ‘2.5D HS’. Being a view-dependent approach, this method is limited to reconstructing only parts of the surface which are directly visible from the virtual camera. This produces holes whenever a partial occlusion is found, as observed in the lateral reconstructions of the arms in the ‘Armadillo’ scene. Furthermore, surfaces which are heavily slanted with respect to the viewing direction will result in gaps in the reconstructed model, as shown in the ‘2.5D HS’ upper right result of the ‘Bunny’ scene. ‘Fused 2.5D HS’ allows to tackle these problems by fusing together different views, but is prone to artefacts where non matching surfaces are united. This is mitigated by the use of Poisson surface reconstruction, which tends to smooth these artefacts, however some of them are still visible in the results.

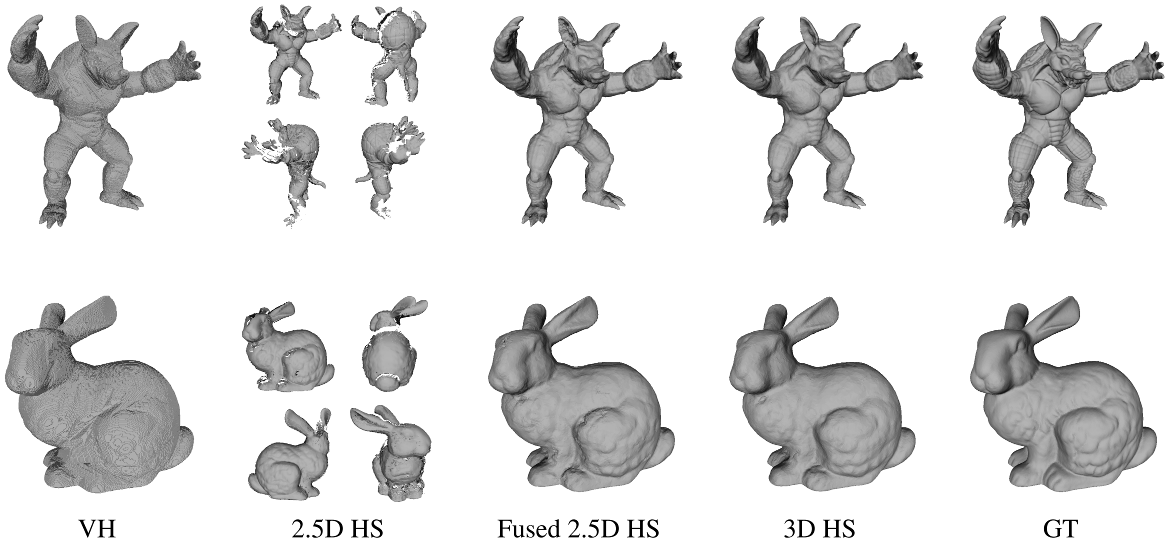


Figure 9: Results from the ‘Armadillo’ and ‘Bunny’ scenes.

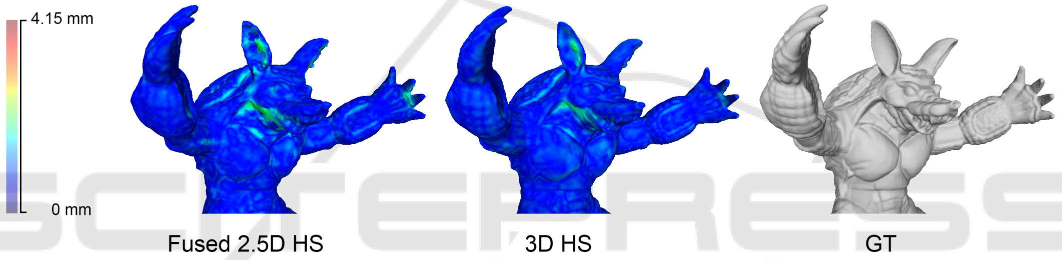


Figure 10: Heatmap showing the accuracy obtained by ‘Fused 2.5D HS’ and ‘3D HS’ when reconstructing the ‘Armadillo’ scene.

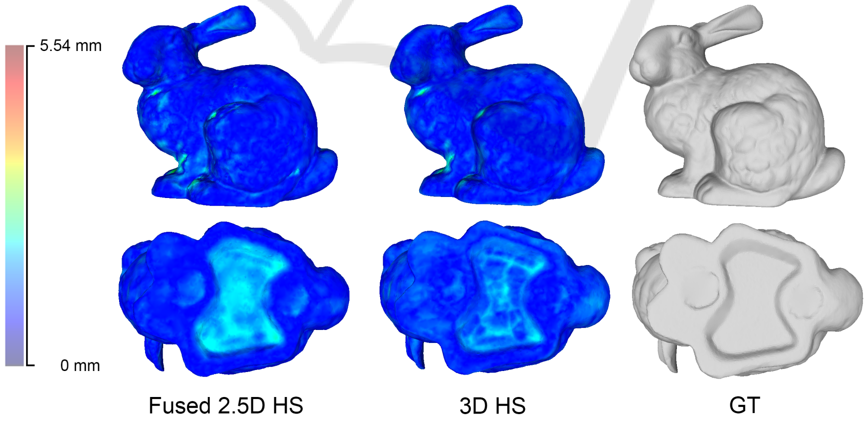


Figure 11: Heatmap showing the accuracy obtained by ‘Fused 2.5D HS’ and ‘3D HS’ when reconstructing the ‘Bunny’ scene.

Finally, ‘3D HS’ is used to perform a full 3D optimisation on top of the previous steps, correcting these artefacts and improving the overall accuracy of the mesh. The results from all the different techniques can be qualitatively compared with the ground truth in the same figure.

In Figure 10 and 11 heatmaps are used to highlight some details of the results where the reconstruction accuracy is improved by ‘3D HS’ with respect to ‘Fused 2.5D HS’. In particular, concavities with a strong error are improved upon by using ‘3D HS’, some notable parts where this can be observed are the ears of the animals in both scenes and the concavity at the base of the ‘Bunny’ scene.

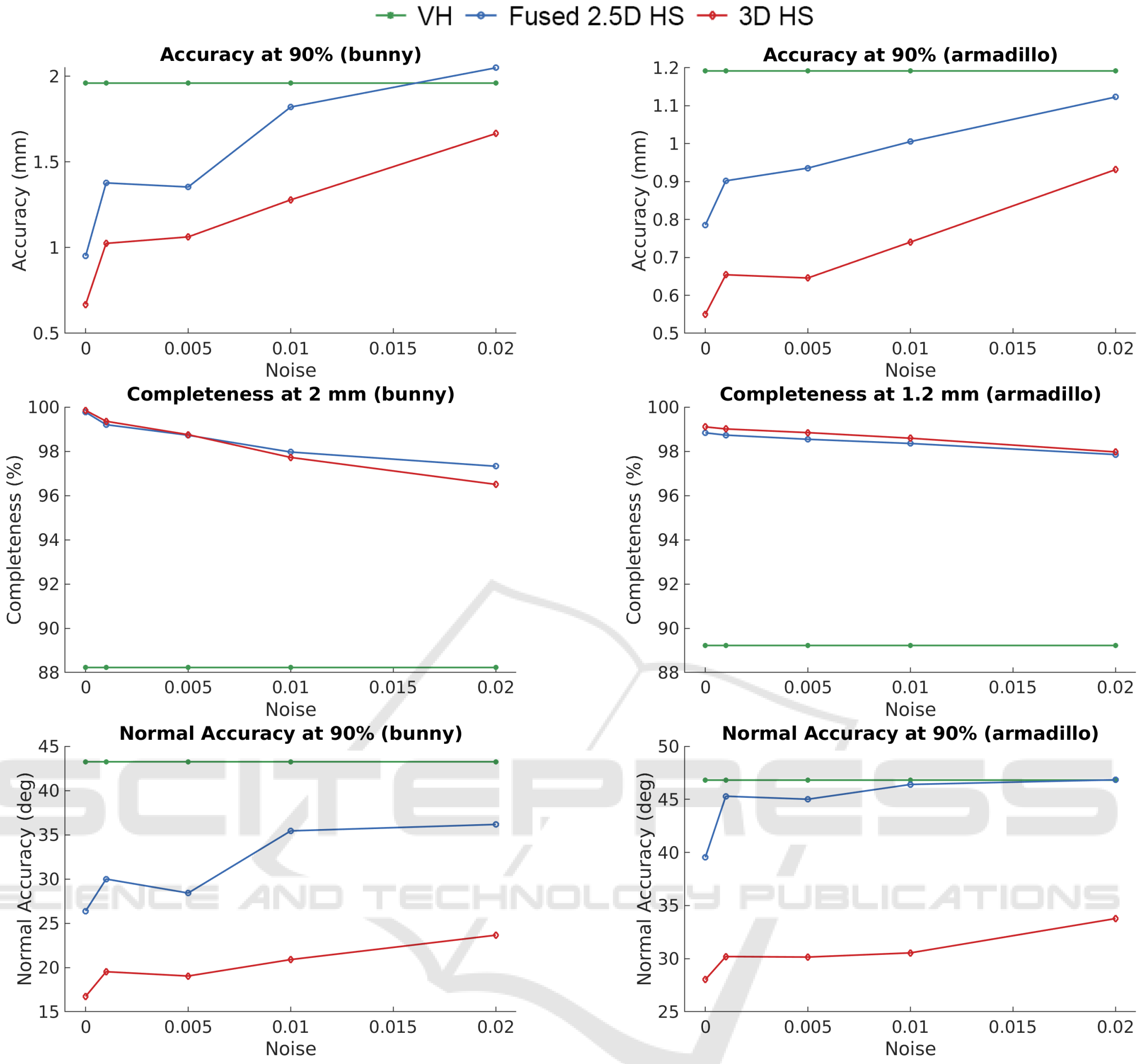


Figure 12: Graphs representing the results on the 'Bunny' and 'Armadillo' scene, including accuracy, completeness and normal accuracy.

Finally, in Figure 12 the performance of these methods is objectively measured at different levels of input noise in terms of accuracy at 90%, normal accuracy at 90% and completeness. As shown in the graphs, '3D HS' achieves sub-millimetre accuracy in both scenes, exceeding 'Fused 2.5D HS' performance by a significant amount. In particular, in the 'Armadillo' scene, which is characterised by a high number of self-occlusions, '3D HS' obtains exceptional results when compared with 'Fused 2.5D HS'. In terms of completeness, both techniques are able to reconstruct the scene properly, without any holes or parts with significant loss of accuracy. In terms of normal accuracy, '3D HS' outperforms 'Fused 2.5D HS', and in particular in the 'Armadillo' scene, which presents high frequency details where it is hard to obtain very precise normals.

The geometric and normal accuracy performance degrades linearly with the introduction of noise, still maintaining good results when the input images are distorted with strong Gaussian noise with a normalised standard deviation of 0.02. This indicates that the approach is robust to noise. In particular, the normal accuracy does not vary significantly, showing how HS normal estimation is robust to noise.

6.2.2 Real Scenes

Figure 13 shows the results obtained in the case of the real scenes. These scene were reconstructed using a very weak initialisation, as can be seen from the 'VH'

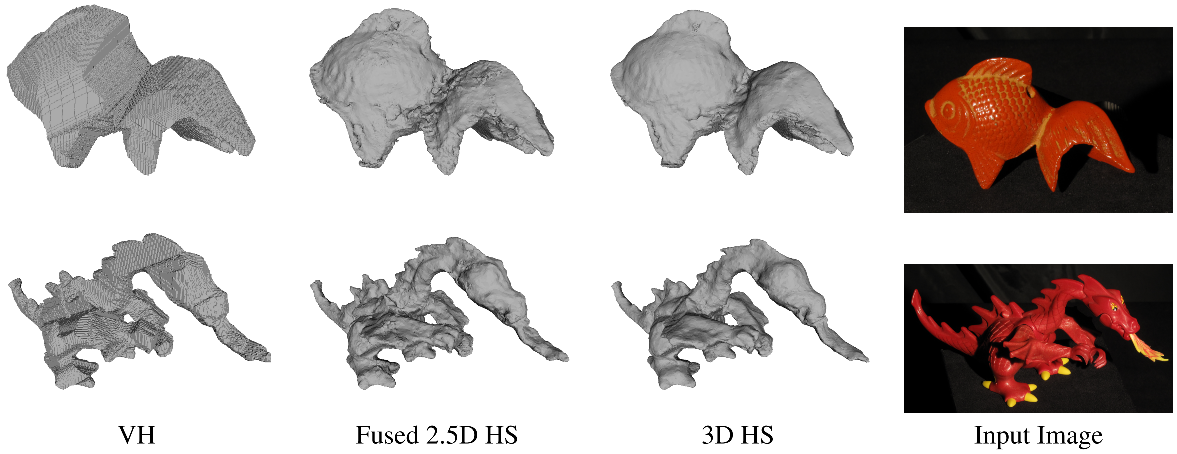


Figure 13: Results from the ‘Fish’ and ‘Dragon’ scenes.

results, due to a lack of precise silhouettes. Starting from this coarse initialisation, the scene is first reconstructed from multiple view points in ‘2.5D HS’, and then the fused depth maps are optimised using ‘3D HS’, leading to accurate models that are able to capture fine structural details including thin structures.

As can be observed from the results, the mesh details improve when performing MRF optimisation on the whole volume, and several artefacts are corrected. It is not possible to perform an objective evaluation for these datasets as no ground truth data is available. However these scenes were included to demonstrate how the methodology is able to work under real conditions. In the ‘Fish’ scene, ‘3D-HS’ is able to obtain fine detail such as the scales despite the presence of strong specularities for this object. In the ‘Dragon’ scene thin structures are correctly reconstructed despite the many self occlusions present in the scene.

7 CONCLUSION AND FUTURE WORK

This paper introduced the first MRF framework for full-3D reconstruction of scenes with unknown complex surface reflectance. The approach proceeds in two steps. First, multiple viewpoint-dependent reconstructions are obtained using a visibility-aware MRF formulation optimised using TRW. The approach uses a VH initialisation to approximate surface visibility and selects the correct cameras to use at each point. The 2.5D surfaces obtained are then fused using Poisson surface reconstruction to obtain an initial full-3D modelling of the scene. Finally, a refined 3D model is obtained through optimisation of a volumetric MRF enforcing Helmholtz reciprocity and normal consistency between neighbouring voxels, before a final mesh representation is extracted using Poisson surface reconstruction. Experimental results demonstrate that the proposed approach is able to achieve sub-millimetre accuracy and significantly outperforms existing 2.5D approaches. Furthermore, the approach has been observed to be robust to high levels of input noise.

Future work will focus on improving the accuracy and efficiency of the approach. In the first stage of the pipeline, viewpoint-dependent reconstruction could be performed from the viewpoint of each camera instead of using a fixed orthographic grid. This would increase scene sampling and allow to take into account the distribution of camera viewpoints. Another interesting avenue for future work would be to avoid using Poisson surface reconstruction, which tends to oversmooth the surface and lead to a loss of detail. This could be achieved by implementing a tailored meshing algorithm which exploits the volumetric representation and preserves normal information estimated through HS. Finally, different MRF optimisation techniques could be investigated to improve reconstruction accuracy. Specific examples of techniques that will be explored include the lazy flipper (Andres et al., 2012), TRW (Kolmogorov, 2006) and higher order cliques approaches (Ishikawa, 2014).

ACKNOWLEDGEMENTS

The authors would like to thank Dr Amael Delaunoy for providing access to the ‘Fish’ and ‘Dragon’ datasets presented in (Delaunoy et al., 2010).

REFERENCES

Andres, B., Kappes, J. H., Beier, T., Köthe, U., and Hamprecht, F. A. (2012). The lazy flipper: Efficient depth-limited exhaustive search in discrete graphical models. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part VII*, ECCV'12, pages 154–166, Berlin, Heidelberg. Springer-Verlag.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48(3):48–259.

Chandraker, M., Bai, J., and Ramamoorthi, R. (2013). On differential photometric reconstruction for unknown, isotropic brdfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2941–2955.

Delaunoy, A., Prados, E., and Belhumeur, P. N. (2010). Towards full 3d helmholtz stereovision algorithms. In *Computer Vision - ACCV 2010 - 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12, 2010, Revised Selected Papers, Part I*, pages 39–52.

Goldman, D. B., Curless, B., Hertzmann, A., and Seitz, S. M. (2010). Shape and spatially-varying brdfs from photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(6):1060–1071.

Guillemaut, J., Drbohlav, O., Illingworth, J., and Sára, R. (2008). A maximum likelihood surface normal estimation algorithm for helmholtz stereopsis. In *VISAPP 2008: Proceedings of the Third International Conference on Computer Vision Theory and Applications, 2008 - Volume 2*, pages 352–359.

Guillemaut, J.-Y., Drbohlav, O., Sra, R., and Illingworth, J. (2004). Helmholtz stereopsis on rough and strongly textured surfaces. In *3DPVT*, pages 10–17. IEEE Computer Society.

Han, T. and Shen, H. (2015). Photometric stereo for general brdfs via reflection sparsity modeling. *IEEE Transactions on Image Processing*, 24(12):4888–4903.

Ishikawa, H. (2014). Higher-order clique reduction without auxiliary variables. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1362–1369.

Janko, Z., Drbohlav, O., and Sara, R. (2004). Radiometric calibration of a helmholtz stereo rig. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–166.

Kazhdan, M., Bolitho, M., and Hoppe, H. (2006). Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, SGP '06, pages 61–70. Eurographics Association.

Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1568–1583.

Krishnamurthy, V. and Levoy, M. (1996). Fitting smooth surfaces to dense polygon meshes. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 313–324.

Laurentini, A. (1994). The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(2):150–162.

Lewis, R. R. (1994). Making shaders more physically plausible. In *In Fourth Eurographics Workshop on Rendering*, pages 47–62.

Liang, C. and Wong, K.-Y. K. (2010). 3d reconstruction using silhouettes from unordered viewpoints. *Image Vision Comput.*, 28(4):579–589.

Lombardi, S. and Nishino, K. (2016). Radiometric scene decomposition: Scene reflectance, illumination, and geometry from RGB-D images. *CoRR*, abs/1604.01354.

Magda, S., Kriegman, D. J., Zickler, T. E., and Belhumeur, P. N. (2001). Beyond lambert: Reconstructing surfaces with arbitrary brdfs. In *ICCV*.

Nasrin, R. and Jabbar, S. (2015). Efficient 3d visual hull reconstruction based on marching cube algorithm. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pages 1–6.

Nishino, K. (2009). Directional statistics brdf model. In *2009 IEEE 12th International Conference on Computer Vision*, pages 476–483.

Oxholm, G. and Nishino, K. (2016). Shape and reflectance estimation in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):376–389.

Roubtsova, N. and Guillemaut, J. (2018). Bayesian helmholtz stereopsis with integrability prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2265–2272.

Roubtsova, N. and Guillemaut, J.-Y. (2017). Colour helmholtz stereopsis for reconstruction of dynamic scenes with arbitrary unknown reflectance. *Int. J. Comput. Vision*, 124(1):18–48.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, CVPR '06, pages 519–528, Washington, DC, USA. IEEE Computer Society.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. (2008). A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE transactions on pattern analysis and machine intelligence*, 30(6):1068–1080.

Tu, P. and Mendonca, P. R. S. (2003). Surface reconstruction via helmholtz reciprocity with a single image pair. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages I–541–I–547 vol.1.

Turk, G. and Levoy, M. (1994). Zippered polygon meshes from range images. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '94, pages 311–318, New York, NY, USA. ACM.

Vogiatzis, G., Hernandez, C., and Cipolla, R. (2006). Reconstruction in the round using photometric normals and silhouettes. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pages 1847–1854, Washington, DC, USA. IEEE Computer Society.

Von Helmholtz, H. and Southall, J. P. (1924). *Helmholtz's treatise on physiological optics, Vol. 1, Trans*. Optical Society of America.

Weinmann, M., Ruiters, R., Osep, A., Schwartz, C., and Klein, R. (2012). Fusing structured light consistency and helmholtz normals for 3d reconstruction. *British Machine Vision Conference*. accepted for publication.

Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):191139–191139–.

Zickler, T. (2006). Reciprocal image features for uncalibrated helmholtz stereopsis. In *IEEE Computer Vision and Pattern Recognitiion*, pages II: 1801–1808.

Zickler, T. E., Belhumeur, P. N., and Kriegman, D. J. (2002). Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision*, 49(2):215–227.

Zickler, T. E., Ho, J., Kriegman, D. J., Ponce, J., and Belhumeur, P. N. (2003). Binocular helmholtz stereopsis. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 1411–1417 vol.2.