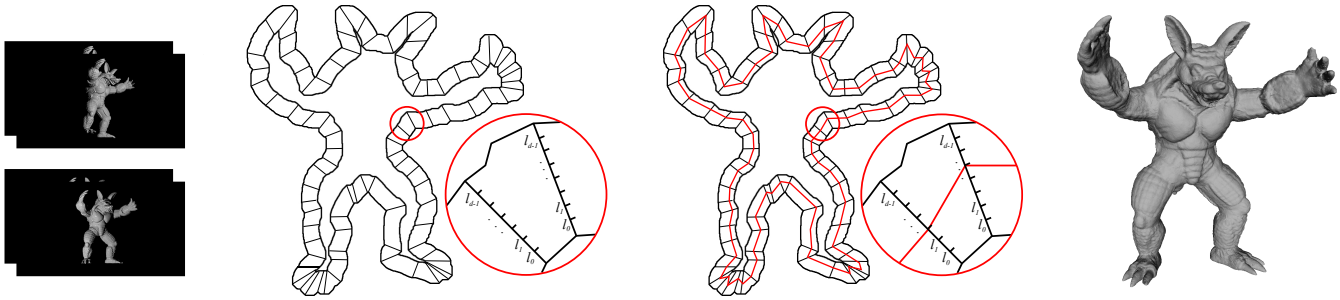
Towards Globally Optimal full 3D reconstruction of scenes with complex reflectance using Helmholtz Stereopsis

Gianmarco Addari and Jean-Yves Guillemaut
Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford, GU2 7XH, UK
{g.addari,j.guillemaut}@surrey.ac.uk



Input Reciprocal Images | Graph Construction | MRF Optimisation | Output Surface

Figure 1: A graphic representation of the proposed method. Helmholtz Stereopsis is used on the input reciprocal images to compute the weights on a Markov Random Field graph. The output surface is obtained by performing global optimisation.

ABSTRACT

Many 3D reconstruction techniques are based on the assumption of prior knowledge of the object's surface reflectance, which severely restricts the scope of scenes that can be reconstructed. In contrast, Helmholtz Stereopsis (HS) employs Helmholtz Reciprocity to compute the scene geometry regardless of its Bidirectional Reflectance Distribution Function (BRDF). Despite this advantage, most HS implementations to date have been limited to 2.5D reconstruction, with the few extensions to full 3D being generally limited to a local refinement due to the nature of the optimisers they rely on. In this paper, we propose a novel approach to full 3D HS based on Markov Random Field (MRF) optimisation. After defining a solution space that contains the surface of the object, the energy function to be minimised is computed based on the HS quality measure and a normal consistency term computed across neighbouring surface points. This new method offers several key advantages with respect to previous work: the optimisation is performed globally instead of locally; a more discriminative energy function is used, allowing for better and faster convergence; a novel visibility handling approach to take advantage of Helmholtz reciprocity is proposed; and surface integration is performed implicitly as part of the optimisation process, thereby avoiding the need for an additional step. The approach is evaluated on both synthetic and real scenes, with an analysis of the sensitivity to input noise performed in the synthetic case. Accurate results are obtained on both types of scenes. Further, experimental results indicate that the proposed approach significantly outperforms previous work in terms of geometric and normal accuracy.

CVMP 2019, Dec. 17–18, London, UK
2019. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS CONCEPTS

• **Computing methodologies** → **Reconstruction**.

KEYWORDS

3D Reconstruction, Helmholtz Stereopsis, Markov Random Fields

ACM Reference Format:
Gianmarco Addari and Jean-Yves Guillemaut. 2019. Towards Globally Optimal full 3D reconstruction of scenes with complex reflectance using Helmholtz Stereopsis. In *Proceedings of the 16th ACM SIGGRAPH European Conference Visual Media Production (CVMP 2019).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

3D reconstruction techniques often rely on simplifying assumptions regarding the scene properties (e.g. surface appearance, its reflectance model or its geometry) [Vogiatzis et al. 2006] or the capture conditions (e.g. camera placement and motion or the illumination conditions) [Han and Shen 2015]. Among those, assumptions on the surface reflectance are central to most reconstruction techniques which usually either assume it follows a particular model or is known a priori. Either type of assumption is problematic. Estimating the specific surface reflectance properties of a scene is very difficult in practice, as it requires estimation of the Bidirectional Reflectance Distribution Function (BRDF) which is high dimensional and hard to compute without dedicated apparatus (e.g. a gonioreflectometer [Nevas et al. 2004; Ward 1992] or structured light projector [Miyashita et al. 2014]). Consequently, restricting the BRDF to follow a specific model is more commonly used, with the Lambertian model being implicit in many algorithms. Under this assumption, any deviations from the model (e.g. specular highlights

The definitive Version of Record was published in European Conference on Visual Media Production (CVMP'19), 2019. DOI: https://doi.org/10.1145/3359998.3369410

in the case of the Lambertian assumption) result in some artefacts or require a dedicated mechanism to reject outliers. Tackling this restriction on surface reflectance is therefore critical to expand the variety of scenes that can be reconstructed and advance the state of the art in 3D reconstruction.

Helmholtz Stereopsis (HS) [Zickler et al. 2002] is a 3D reconstruction technique that is agnostic to the scenes' reflectance properties. It is based on Helmholtz reciprocity [Von Helmholtz and Southall 1924], which consists in the formulation of a constraint on the surface normal that is independent of the BRDF. Work on this technique has been mostly limited to partial scene reconstruction (i.e. 2.5D reconstruction [Guillemaut et al. 2008; Roubtsova and Guillemaut 2018]), while recent extensions to full 3D reconstruction fail to guarantee global optimality [Delaunoy et al. 2010]. Furthermore, visibility is not handled in 2.5D reconstruction formulations [Roubtsova and Guillemaut 2018] as all the cameras are confined to one side of the object to avoid occlusions, while in 3D formulations [Addari and Guillemaut 2019] it is computed approximately. This is particularly problematic for generalisations to full 3D scenes which may present severe occlusions and require a mechanism to explicitly handle it.

In this paper, a novel approach based on HS is proposed together with a technique to handle visibility in HS datasets. A graphic representation of the proposed pipeline is shown in Figure 1. The methodology consists in estimating the target surface by optimising a Markov Random Field (MRF) based on Helmholtz saliency and normal consistency. The approach only requires a coarse initialisation to define the search space; this may be provided by a Visual Hull (VH) or a rough reconstruction obtained from another method. From this initialisation, two surfaces are found to reasonably bind the search space and construct the MRF graph. An energy function is then defined based on the quality and consistency of the normals estimated using HS. As the energy function used during optimisation is nonsubmodular, global optimality cannot be fully guaranteed, however using a Tree Reweighted Message Passing (TRW) algorithm allows to obtain a high quality solution with partial guarantees on optimality [Kolmogorov 2013]. A new visibility estimation approach is also proposed to tackle the lack of geometric information on the scene. The method is based on selecting the camera pairs that produce a surface normal with a strong confidence measure for the hypothesised surface points. The proposed approach is then evaluated qualitatively and objectively against state-of-the-art results, significantly outperforming them in terms of geometric and normal accuracy.

The rest of the paper is structured as follows. The next section surveys the main 3D reconstruction approaches, highlighting their dependency on the scene's reflectance properties as well as the state of the art in HS and its underpinning principles. Then, the proposed methodologies are introduced, detailing the key concepts of the reconstruction approach and the new visibility handling procedure used to enhance the results. A thorough evaluation follows, comprising a qualitative analysis on both real and synthetic scenes and quantitative assessment on the synthetic datasets. Finally, the conclusions of this work are drawn and possible future directions are discussed.

2 RELATED WORK

2.1 3D Reconstruction Techniques

The most common 3D reconstruction techniques can be broadly divided in three categories: Shape from Silhouettes (SfS), Multi-View Stereo (MVS) and Photometric Stereo (PS).

SfS general principle consists in intersecting the set of visual cones defined by backprojecting the object's silhouette in each image [Laurentini 1994; Liang and Wong 2010; Nasrin and Jabbar 2015]. Independence from the surface properties is the main advantage offered by this class of methods as long as good background segmentation can be obtained. They are however inherently limited as they only allow reconstruction of convex objects due to the impossibility of visualising concavities in silhouettes.

Binocular and Multi-view Stereo methods [Seitz et al. 2006; Szeliski et al. 2008] use point correspondences across images to infer surface depth. Contrary to SfS they are not limited to the reconstruction of convex objects. Nonetheless, the reconstructed surfaces are often assumed to be Lambertian or sufficiently textured to perform point matching across different views. This assumption is the main drawback to classic stereo approaches as it is often violated. In [Lombardi and Nishino 2016; Oxholm and Nishino 2014] surface reflectance is jointly estimated with scene geometry. The technique consists in iteratively refining one term at a time by relying on the previous estimate of the other term. This hinders the results greatly, as each solution will only be as good as the previously estimated other term.

A more reliable technique for surfaces with complex reflectance is Photometric Stereo (PS) [Woodham 1980]. This family of approaches allows for reconstruction of non-Lambertian surfaces, however the surface reflectance needs to be known a priori. There are some exceptions such as [Chandraker et al. 2013], where image gradients are used to perform surface reconstruction with unknown reflectance. This method is however limited to isotropic BRDF. In [Goldman et al. 2010], a set of predetermined BRDFs are used as a basis and mixed to estimate surface reflectance and perform geometric reconstruction, while in [Han and Shen 2015] a complex set-up is presented to provide a very densely set of light and viewing directions. This allows to exploit specularities and shadows produced to obtain the final solution.

In recent years Convolutional Neural Networks (CNNs) have become ubiquitous in several computer vision applications. They offer an important contribution to 3D reconstruction as they allow for single-view scene computation. However, work in this area is still constrained by the training data. In [Kar et al. 2015], the proposed approach is solely able to perform reconstruction on objects from a specific set of categories. In [Choy et al. 2016], the limitation of object labels is overcome, however the results consist in low resolution volumetric renditions of the analysed scenes. Another contribution, proposed in [Firman et al. 2016], consists in a probabilistic volumetric scene completion. Despite the potential of CNNs based methods, they have yet to reach the level of accuracy of other types of approaches when a larger number of images are used.

None of these methods fully solve the problem for surfaces presenting complex and varying reflectance. They mainly rely on simplifying assumptions to constrain the problem. An alternative to

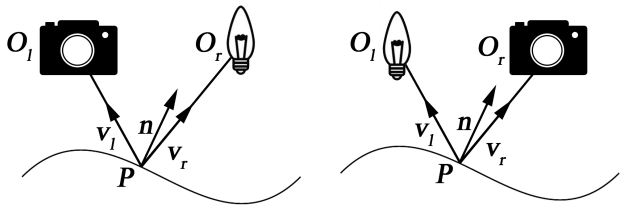


Figure 2: Base HS set-up. A reciprocal pair of pictures is taken by inverting camera and light positions. The surface point BRDF is invariant in the two scenarios.

these techniques that allows to reconstruct surfaces with complex BRDFs is HS.

2.2 Helmholtz Stereopsis

HS was first proposed by Zickler et al in [Zickler et al. 2002]. It consists in using Helmholtz reciprocity [Von Helmholtz and Southall 1924] as a constraint to identify 3D points located on the surface of an object and their corresponding normals. Helmholtz reciprocity states that the measured BRDF at a surface point will stay constant when inverting emittent and receiving directions. This BRDF invariance is exploited to compute surface information using multiple pairs of images from which the point is visible. In its first formulation, HS utilises a maximum likelihood formulation to find the object's surface, and no regularisation is performed to smooth the obtained result, which can yield noisy and inconsistent solutions.

Before reviewing more recent work on HS, it is important to introduce its main concepts, which will be referred to in the methodology section. A representation of the base set-up is shown in Figure 2. From this representation, it can be observed that given a unit-strength point light source located at $\boldsymbol{O_r}$, the intensity of the projection of point $\boldsymbol{P}$ measured at the viewing position $\boldsymbol{O_l}$ will be:

$$i_l = f(\boldsymbol{v}_r, \boldsymbol{v}_l) \cdot \frac{\boldsymbol{n} \cdot \boldsymbol{v}_r}{|\boldsymbol{O_r} - \boldsymbol{P}|^2} \tag{1}$$

where $f(\boldsymbol{v}_r, \boldsymbol{v}_l)$ is the BRDF at point $\boldsymbol{P}$. Unit vectors $\boldsymbol{v}_r$ and $\boldsymbol{v}_l$ respectively indicate lighting and viewing directions, while $\boldsymbol{n}$ is the surface normal at point $\boldsymbol{P}$. When switching camera and light position, an equivalent equation can be found for the projection of $\boldsymbol{P}$ at the viewing point $\boldsymbol{O_r}$.

As mentioned, Helmholtz reciprocity states that the BRDF of a surface point remains invariant when inverting viewing and emittent directions. In this set-up, this property translates to $f(\boldsymbol{v}_r, \boldsymbol{v}_l) = f(\boldsymbol{v}_l, \boldsymbol{v}_r)$. By applying substitution, the following can be obtained:

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

which indicates that vector $\boldsymbol{w}$ will be tangent to the surface and is independent from the BRDF at point $\boldsymbol{P}$.

By computing multiple vectors $\boldsymbol{w}$ from a series of reciprocal camera light pairs for a fixed point $\boldsymbol{P}$, the normal $\boldsymbol{n}$ can be obtained as the vector perpendicular to all instances of $\boldsymbol{w}$. Moreover, given multiple hypotheses for the position of $\boldsymbol{P}$, it is possible to compute a confidence measure for each position, by measuring how reliable the normal estimation is at each instance of $\boldsymbol{P}$. In particular, if the vectors $\boldsymbol{w}$ are grouped as rows of a matrix, called W, both the normal estimate and its confidence score can be obtained by performing Singular Value Decomposition (SVD):

$$SVD(W) = U\Sigma V^T \tag{3}$$

the normal will be the last column of V, while the confidence measure can be obtained from the non-zero components of diagonal matrix Σ.

Initial work on HS includes [Zickler et al. 2003] and [Tu and Mendonca 2003], where a single pair of reciprocal images is used to perform HS. This is made possible by constraining the reconstructed surface to have C_1 continuity. In [Janko et al. 2004] radiometric calibration is proposed to significantly improve normal estimation through HS, while in [Guillemaut et al. 2004] the authors extend the surfaces that can be reconstructed to strongly textured and rough surfaces by computing HS on image patches instead of using single pixel back-projection. Radiometric calibration is then proposed in a different form in [Zickler 2006], where it is performed through the observation of specular highlights across the Helmholtz image pairs. Finally, in [Guillemaut et al. 2008] a different confidence measure is proposed in the form of the radiometric distance function, which considerably improves previous results using maximum likelihood.

An alternative to maximum likelihood for classic HS is first proposed in [Roubtsova and Guillemaut 2017] and [Roubtsova and Guillemaut 2018], where a maximum a posteriori formulation is proposed in both white light HS and in a new approach named colour HS, where the use of coloured lights paired with filtered cameras allows to perform dynamic surface reconstruction. All the aforementioned techniques are applied to 2.5D surfaces, and do not handle visibility or occlusions in any way.

The first time HS is used to perform full 3D reconstruction is in [Delaunoy et al. 2010], where gradient descent is proposed to perform the optimisation. The faces of the initial surface are iteratively moved towards a lower energy solution. This method is prone to optimising towards local minima, unless the initialisation is chosen very close to the target surface. Two other approaches were then proposed in [Addari and Guillemaut 2019], the first one consists in fusing together multiple 2.5D surfaces taken from different points of view, while the second one is based on volumetric optimisation and uses a multi-labelled MRF to find the point in each voxel where the surface is most likely to cross. Both methods use a maximum a posteriori optimisation to produce their results, however global optimality is not guaranteed in the first approach, while the second approach fails to provide a good regularising term in its energy function and is also impractical as it relies on Iterative Conditional Modes (ICM) in its implementation, an exhaustive search optimiser.

3 METHODOLOGY

The key idea of the proposed methodology is to introduce a one step approach to perform full 3D reconstruction from a coarse initialisation, providing a mechanism to perform global optimisation and recover a solution with strong optimality properties through the use of state-of-the-art MRF solvers. The target surface is obtained through a maximum a posteriori approach using an MRF graph. A new technique to handle visibility in HS datasets is jointly

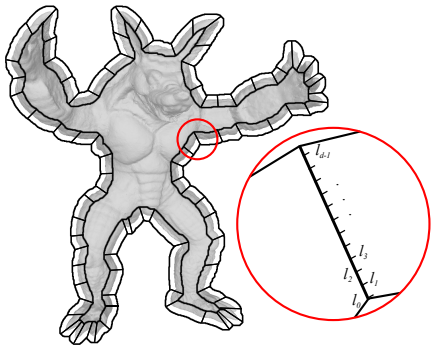


Figure 3: The labelling indicates 3D points positioned at regular intervals between corresponding vertices from the outer and inner surfaces.

proposed, based on selecting a small number of camera pairs that produce the best HS confidence measure among all views.

The 3D reconstruction is performed through a pipeline that is exemplified in Figure 1. The first step is to define the solution space by producing two surfaces that contain the surface of the object. The outer surface must encompass the object fully, although it is not required to be a precise fit; while the inner surface must be completely inside the surface of the object. Correspondences are drawn between points on the two surfaces by performing surface registration and are used to construct a multi labelled MRF graph. Each pair of registered vertices from the two surfaces defines a node in the graph, while the labelling corresponds to regularly spaced intermediate points between the two vertices. In order to perform the optimisation, an energy function is defined, based on an HS confidence measure and normal consistency across the surface. By minimising said energy function, the target surface is obtained.

3.1 Graph Construction

To initialise the method and identify the search space over which the optimisation is performed, two surfaces are defined. The first one corresponds to the outer boundary of the solution space and must completely encompass the object. For instance, the VH of the object or an accordingly dilated approximation, obtained from a different technique, could be used. The second surface must instead be completely inside the object, while maintaining a similar topology to the outer surface; this can be achieved by carving the outer surface.

In order to draw correspondences between the two surfaces, the outer one is registered onto the inner one by using non-rigid Iterative Closest Point (ICP) [Audenaert et al. 2019]. The use of a non-rigid approach allows to match key features of the surfaces despite their difference in scale. The result is a dense matching between the two surfaces, where a point of the target surface is necessarily found between each pair of registered vertices.

These correspondences are then used to construct an MRF graph. Each pair of corresponding vertices between the two surfaces will correspond to a node and its neighbours will be established depending on the surface topology. In practice, the surface is represented as a 3D mesh from which each edge corresponds to a graph edge, defining the neighbourhood of the two bound vertices. Each node is then assigned a set of labels $\{l_0, l_1, ...l_{d-1}\}$, where each label indicates a 3D point on the segment connecting the two surfaces at the node's corresponding vertices. A labelling example is illustrated in Figure 3. In particular, l_0 will coincide with the vertex on the outer surface, l_{d-1} with the corresponding vertex on the inner surface and all the intermediate labels will be spaced regularly in between them. In the following section, the energy function used to perform the optimisation is detailed.

3.2 Energy Formulation

To reconstruct the target surface an energy function is minimised. The function is composed of a data term, that expresses the likelihood of finding the surface at a certain 3D point, and a smoothness term which allows for a regular surface to be produced. The aim of the optimisation process is to compute the label for each node where the target surface is most likely to intersect the segment. The 3D points defined by the labelling and the connections already established between them are then used to reconstruct the output surface.

The energy function is defined as follows:

$$E(L) = (1-\alpha)\sum_{v\in\mathcal{M}} D(X(v,l_v)) + \alpha \sum_{v,w\in\mathcal{N}_M} S(X(v,l_v),X(w,l_w)) \quad (4)$$

where L indicates the whole set of labels across all nodes, α is a weighting factor to balance the effect of data and smoothness terms and $\mathcal{N}_M$ defines a property for which two vertices are neighbours on the surface $\mathcal{M}$. D and S are respectively the data and smoothness terms, calculated for all the nodes and edges of the graph; while the operator $X(v,l_v)$ is used to identify node v 3D position when it is assigned the label l_v.

The data term is based on HS confidence measure, and computed in a similar fashion to [Addari and Guillemaut 2019; Roubtsova and Guillemaut 2018], with an additional term to prevent assigning labels located outside the VH:

$$D(\boldsymbol{P}) = \begin{cases} \infty, & \text{if } \boldsymbol{P} \notin VH \\ 1, & \text{if } |vis(\boldsymbol{P})| < 3 \\ e^{-\mu \times \frac{\sigma_2(\boldsymbol{P})}{\sigma_3(\boldsymbol{P})}}, & \text{otherwise} \end{cases} \quad (5)$$

where σ_2 and σ_3 indicate the second and third diagonal terms of matrix Σ as computed in Equation 3 and $vis(\boldsymbol{P})$ is the set of camera pairs from which point $\boldsymbol{P}$ is visible. Points that are found to be outside the VH are assigned an infinite weight, since the VH must contain the object; while points that are not visible from enough cameras are given a strong weight that still allows for self occlusions to be reconstructed. The way visibility is handled in this approach is explained in Section 3.3.

The smoothness term serves as a regularising weight to ensure the surface is smooth and consistent with the photometric normals calculated through HS. It is based on a depth disparity measure, here referred to as $\delta(\boldsymbol{V}, \boldsymbol{W})$, calculated between pairs of neighbouring nodes. $\delta(\boldsymbol{V}, \boldsymbol{W})$ consists in the distance between a point and its

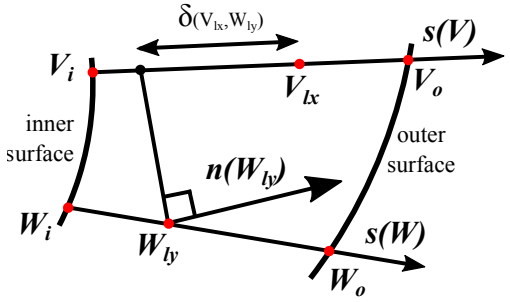


Figure 4: Graphic representation of how the smoothness term is computed. The point W_{l_y} is projected perpendicularly to its estimated normal towards the segment $V_i V_o$ and the error is measured as the distance between the projection and point V_{l_x}.

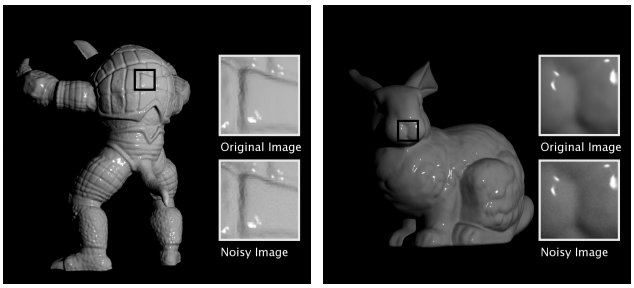


Figure 5: Armadillo and Bunny dataset image with input noise. The inset images show a comparison between the original image and the one with input noise.

predicted position based on the estimated normal of its neighbour. It is calculated as follows:

$$\delta(\boldsymbol{V}, \boldsymbol{W}) = \frac{\boldsymbol{VW} \cdot \boldsymbol{n(W)}}{\boldsymbol{n(W)} \cdot \boldsymbol{s(V)}} \tag{6}$$

where $\boldsymbol{VW}$ indicates the vector connecting the two points and $\boldsymbol{s(V)}$ is a unit vector representing the direction of the segment that connects inner and outer surface at the node corresponding to point $\boldsymbol{V}$. The disparity error is computed as the difference between point $\boldsymbol{V}$ and the projection of $\boldsymbol{W}$ perpendicular to its estimated normal $\boldsymbol{n(W)}$ towards said segment. In Figure 4 it is shown a simplified representation of the smoothness term calculation.

This term is a generalisation to a perspective sampling in full 3D of the depth disparity measure presented in [Roubtsova and Guillemaut 2018]. Moreover, the error measure here presented is more discriminative than the one presented in [Addari and Guillemaut 2019], which is tied to the voxel size chosen and where strong discontinuities do not result in a considerable error. In contrast, the proposed distance penalises more heavily depth and normal assignments which are inconsistent between neighbouring nodes, being not bounded by the volume sampling resolution.

The smoothness term is then computed as the average of the squared disparity terms for the two neighbours, and it is truncated at a threshold of t^2:

$$S(\boldsymbol{V}, \boldsymbol{W}) = \begin{cases} \frac{1}{2}(\delta(\boldsymbol{V}, \boldsymbol{W})^2 + \delta(\boldsymbol{W}, \boldsymbol{V})^2), & \text{if } \delta(\boldsymbol{V}, \boldsymbol{W}) \text{ and } \delta(\boldsymbol{W}, \boldsymbol{V}) < t \\ t^2, & \text{otherwise} \end{cases} \tag{7}$$

The threshold is used to allow for natural discontinuities, and it is also used where occlusions do not allow HS to produce an estimated normal.

3.3 Visibility Computation

A probabilistic approach based on HS is here proposed to perform visibility computation. In [Addari and Guillemaut 2019], visibility is computed by approximating each point to the closest position on the initialisation surface, and then using the same surface to find occlusions. Similarly, in [Delaunoy et al. 2010] the surface obtained at each iteration is used to compute visibility. In both cases, depending on the initialisation used or current solution, visibility estimation is not robust and may consider cameras from which the point is actually occluded. To avoid this, the following approach is proposed.

A first selection of cameras is performed by approximating all labels to their corresponding point on the outer surface, and computing occlusions for said point. A further selection is then performed on these cameras by finding the k pairs that have the highest likelihood of producing a good normal. In this paper, k was chosen to be three since this is the minimum number of camera pairs needed to perform HS. To do so, k pairs of the remaining cameras are iteratively selected randomly and the confidence measure is computed for their corresponding W matrix obtained as shown in Equation 3. The subset that minimises the following equation is then chosen:

$$\min_{c_0,\ldots,c_{k-1}} \frac{\sigma_3(W_{c_0,\ldots,c_{k-1}})}{\sigma_2(W_{c_0,\ldots,c_{k-1}})} \tag{8}$$

These cameras have a high likelihood of having full view of the current point as they will have a strong agreement over the point normal. Furthermore, selecting a fixed number of cameras for each point provides the additional advantage of a more consistent confidence score when computing the data term than using different numbers of cameras for each point and speeds up computation.

3.4 MRF Optimisation

The final aspect taken into consideration in this methodology is the technique that can be used to perform the final optimisation. The energy function chosen to represent the problem violates the submodularity constraint, and non submodular functions cannot be properly minimised by classic graph-cut approaches as indicated in [Kolmogorov and Rother 2007]. However, many techniques exist to approximate the solution of a nonsubmodular function with a high degree of confidence. The approach chosen in this paper is TRW [Wainwright et al. 2005] in its more recent formulation called Sequential Tree Reweighted Message Passing (TRW-S) [Kolmogorov 2006, 2013], which contrary to TRW guarantees that the energy lower bound does not decrease during optimisation and introduces the condition of weak tree agreement to identify local maxima in the energy bound.

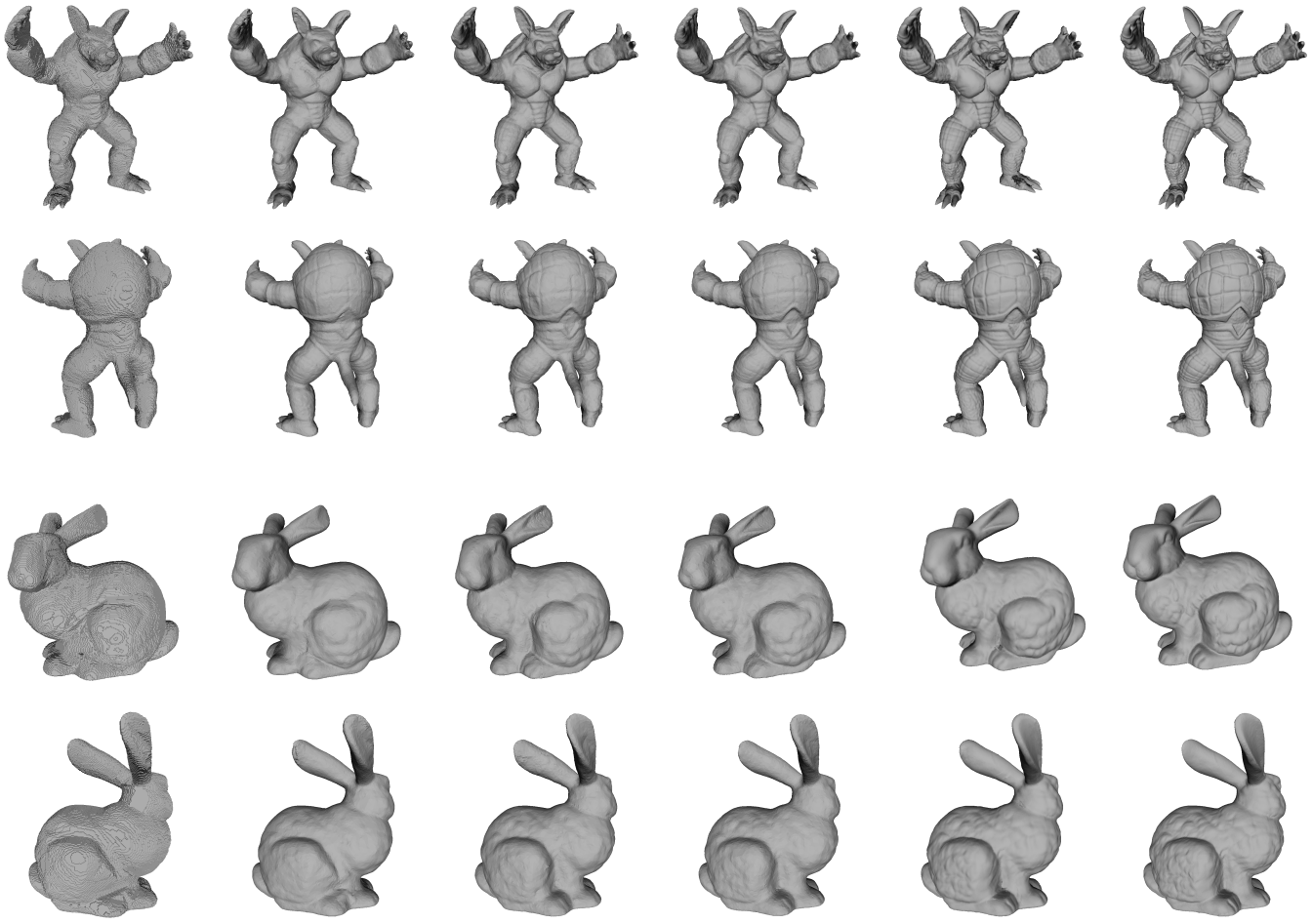


Figure 6: Results obtained on the synthetic scenes from the evaluated methods.

4 EXPERIMENTAL EVALUATION

In this section the proposed approach is evaluated against the following methods: 'VH' based on SfS; 'Fused 2.5D HS' and '3D HS' from [Addari and Guillemaut 2019] and an additional method based on graph-cut which will be indicated as 'Graph-Cut HS'. This method is inspired by voxel carving from [Vogiatzis et al. 2007], and it consists in estimating the voxel occupancy in a grid by performing graph-cut on a binary labelled MRF. It is included here as it provides a baseline, being a classical volumetric approach to 3D reconstruction which we adapted to HS for comparison purposes. The data term utilised here is the constant ballooning term and it is only applied to outside voxels in order to avoid an empty solution. The smoothness term is applied at the edges of neighbouring voxels and is based on the HS confidence measure. In particular, it is computed as follows:

$$G(V,W) = \begin{cases} e^{-\mu \times \frac{\sigma_2(\boldsymbol{P})}{\sigma_3(\boldsymbol{P})}}, & \text{if } l_V \neq l_W \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where V and W are neighbouring voxels, $\boldsymbol{P}$ is the central point at the boundary between the two voxels and l_V indicates the labelling of the node corresponding to V. In this method, the HS quality measure is used as a regularising term for the surface and thus calculated at the boundary. The optimisation is then performed to compute the voxel occupancy and the 3D points at the boundary are used together with the HS estimated normals to perform surface integration.

4.1 Synthetic scenes

To produce a quantitative evaluation of the methods, the datasets from [Addari and Guillemaut 2019] were used. They include two synthetic scenes (Armadillo and Bunny) rendered with a physically plausible engine so that Helmholtz reciprocity is respected. They are both characterised by a specular BRDF. The dataset consists of 40 pairs of views acquired using cameras regularly sampled on a sphere around the objects. All images were captured at a 1920×1080 resolution. For these scenes the analysis was performed by altering the images with Gaussian noise to emulate a more realistic scenario.

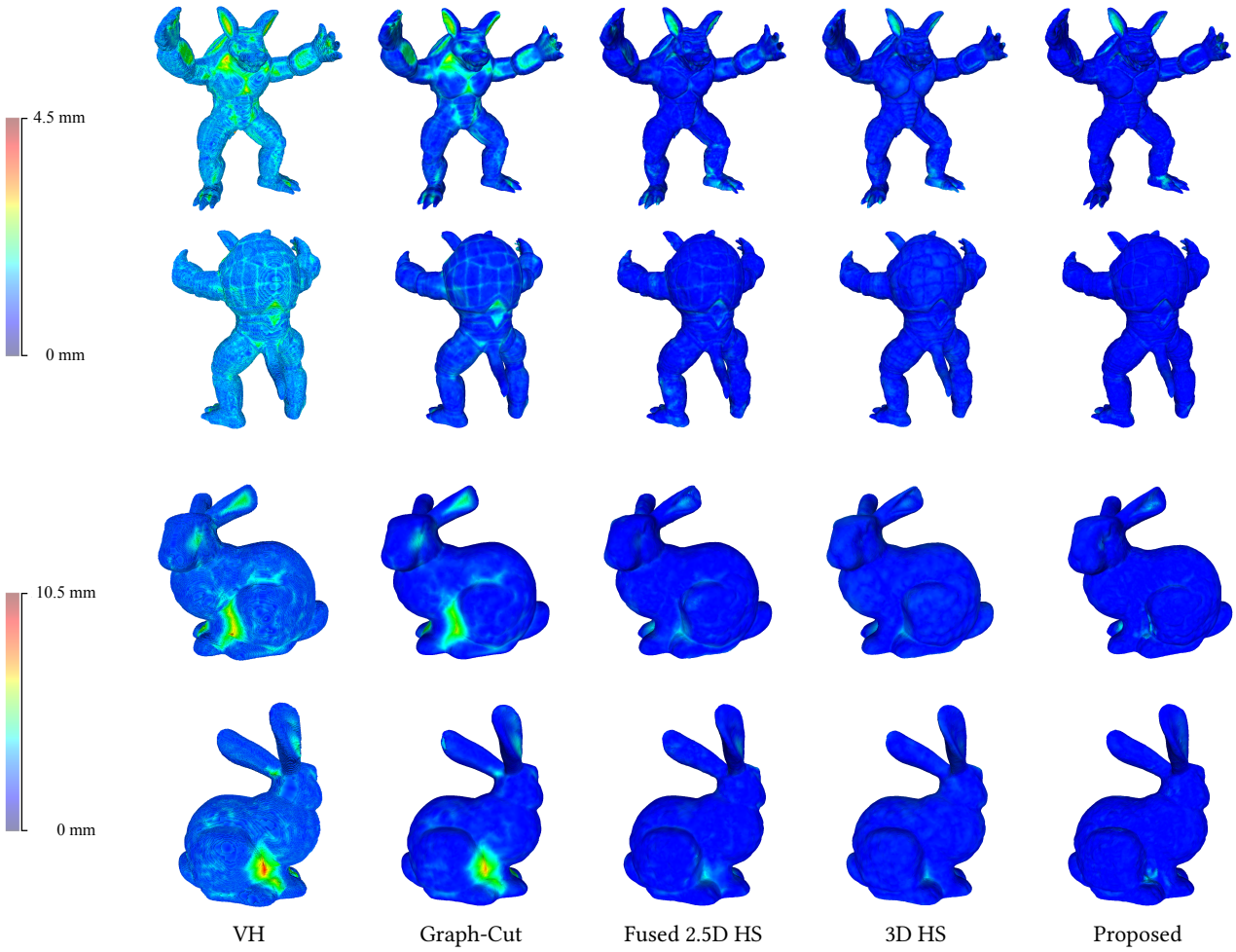
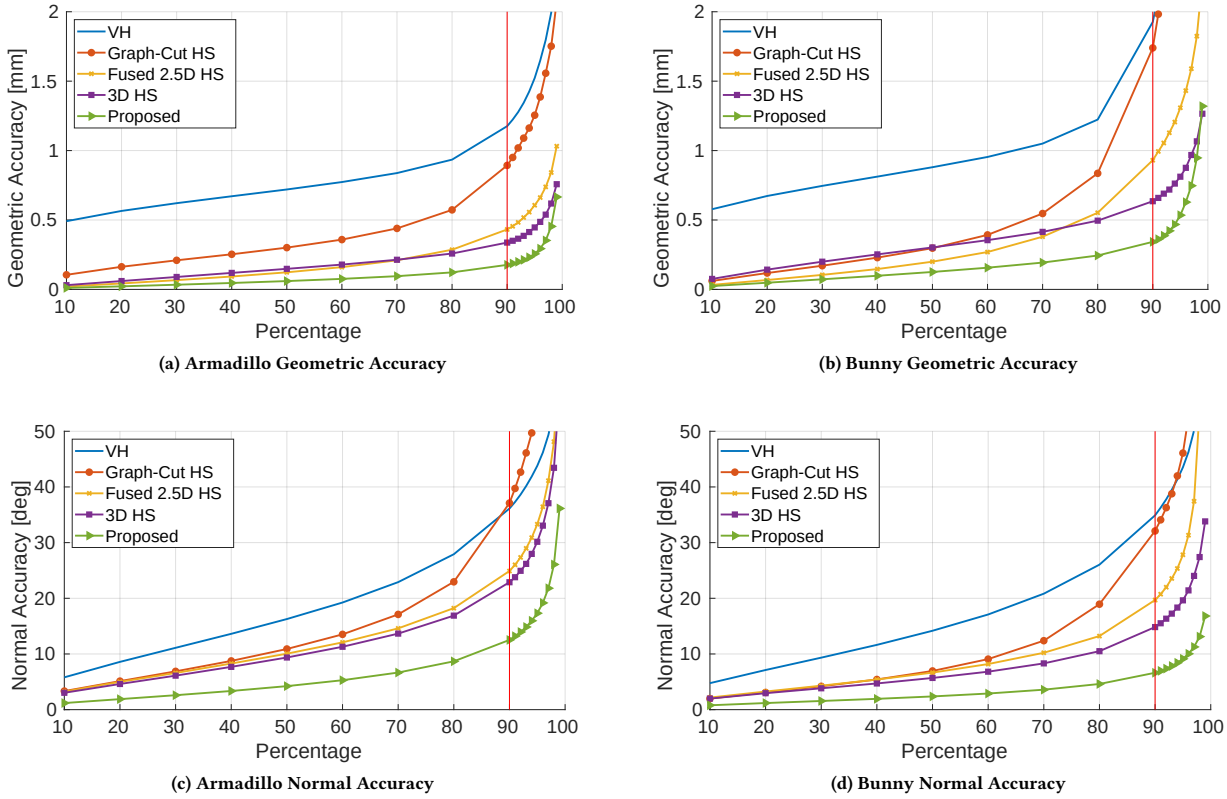


Figure 7: Error Maps for the synthetic scenes computed using Hausdorff distance.

In particular, a standard deviation of 0.1% of the full 16 bit image range was used. The effect on the images can be observed in the inset images in Figure 5.

The results for the synthetic scenes are presented side by side with the Ground Truth (GT) in Figure 6. As it can be observed, the 'VH' method performs poorly, especially with regards to reconstructing concavities. This is clearly shown by the lack of details in both scenes. In comparison, the additional method 'Graph-Cut' is able to recover more surface details due to the use of HS for normal estimation. This method, however, shows the limitations of relying on conventional approaches such as the ballooning term to regularise the HS saliency measure. The use of a separate step to perform surface integration contributes to oversmoothing the final results of this method. 'Fused 2.5D HS' achieves better results with respect to the previous methods by introducing a tailored regularisation term enforcing the consistency of the depth and normal estimates. Nonetheless, some artefacts are present on the surface due to the fusion of the separately computed partial surfaces. While this is somewhat mitigated by the use of Poisson surface reconstruction, fine details present on the surfaces are lost in the process. The results obtained by '3D HS' overcome some of these limitations by introducing a volumetric optimisation step on top of the previous results. The use of ICM during optimisation is however iterative and ultimately unable to fully retrieve minute concavities present on the surfaces. As with the previous methods, '3D HS' is also further hindered by the use of Poisson surface reconstruction after the optimisation is performed. The proposed method achieves the complete retrieval of fine details from the surfaces such as the shell grooves of the Armadillo and the facial features of the Bunny.

In Figure 7 a further comparative analysis can be done by observing the Hausdorff distance on specific areas of the scenes with respect to the GT. In particular, it can be noted how critical areas such as the shell of the Armadillo present minimal error in the proposed approach as opposed to all other methods where its concavities are not accurately reconstructed. Whereas in the Bunny scene the main improvement can be observed in the ears, where the error is significantly lessened.



(a) Armadillo Geometric Accuracy

(b) Bunny Geometric Accuracy

(c) Armadillo Normal Accuracy

(d) Bunny Normal Accuracy

Figure 8: Geometric and Normal Accuracy graphs for the synthetic scenes.

To conclude the analysis on the synthetic scenes, we compute the Middlebury geometric and normal accuracy [Seitz et al. 2006] for all methods shown in Figure 8. These measures are performed for each vertex in the resulting surfaces by selecting the closest point on the ground truth and measuring the geometric and normal differences. At the top, the scenes are analysed in terms of geometric accuracy on all methods, while at the bottom the graphs show the normal accuracy achieved. The graphs show how the geometric and normal accuracies vary for each method as a wider percentage of the surface is considered (shown on the horizontal axis). It can be observed that the proposed method significantly outperforms all other approaches by achieving a geometric accuracy at 90% of 0.17 mm and 0.32 mm respectively in the Armadillo and the Bunny scenes. The normal accuracy is also significantly improved due to the global approach to optimisation and the implicit performance of surface integration during the MRF graph construction.

It must be noted that the results for the proposed approach presented in Figure 6 were rendered using the Helmholtz normals estimated during optimisation. Performing the surface integration jointly with the optimisation offers the unique advantage of preserving the photometric normals information at the same 3D locations of all vertices in the final surface. An example of the difference between performing flat shading and using the additional normal information is shown in Figure 9.

4.2 Real scenes

To evaluate the approaches on real scenes, the dataset from [Delaunoy et al. 2010] was used. The two scenes (Fish and Dragon) were chosen due to the strong specularities of the fish material and the many self occlusions of the dragon. They are respectively comprised of 72 and 68 views acquired using a turn-table at a resolution of 1104×828.

In Figure 10 the results obtained on these scenes from each method are compared. As it can be appreciated, the 'VH' reconstructions are particularly rough. This is due to a poor segmentation of the data, which is especially challenging to obtain properly in HS setups because of the change in lighting across different views. The 'Graph-Cut' method fails to obtain a closed surface due to the lack of views from the bottom side of the object, which results in some artefacts on the lower part of the scenes. The 'Fused 2.5D HS' results are extremely noisy due to the superposition of multiple 2.5D surfaces that are incoherent. This can be observed especially at the base of the tail fin of the Fish. The artefacts produced by the fusion is partially mitigated in the '3D HS' results, however the use of Poisson surface reconstruction tends to smooth out some finer details of the scene. Finally, the results from the proposed method present an accurate reproduction of most minute aspects of the object. In the Fish scene, the small cavity inside the eye is maintained as well as the features on the side of the tail fin and the two small

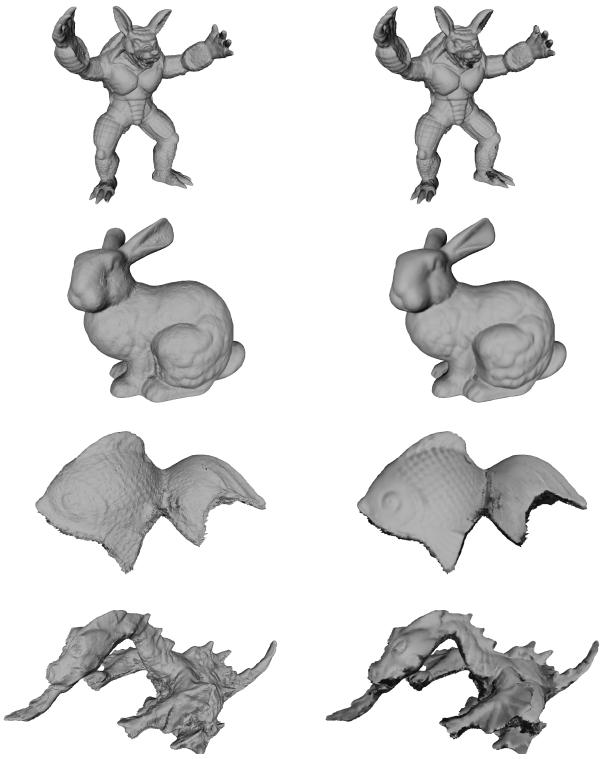


Figure 9: Comparison between using Flat Shading and Smooth Shading using the estimated Helmholtz Normals.

concavities behind the dorsal fin. The high-frequency details of the scales on the side of the object are also faithfully reconstructed, while they are oversmoothed in the previous approaches due to poor surface integration. In the Dragon scene, the scales motif on the side is accurately reconstructed and more details can be seen on the face with respect to what previous approaches achieved.

To summarise, the proposed approach produces significantly more accurate results in comparison with the previous methods considered. The disadvantages of a weak initialisation are overcome with the use of the visibility handling approach, which also allows to better handle occlusions.

5 CONCLUSIONS AND FUTURE WORK

In this paper, a BRDF independent approach for full 3D reconstruction was proposed. The method consists in a maximum a posteriori optimisation based on HS. The method is initialised by defining a solution space that fully encompasses the scene, represented as an outer and inner surface. An MRF graph is then constructed inside the solution space by assigning a node to each corresponding pair of points on the surfaces; while the labelling indicates 3D points in between corresponding nodes. An energy function based on HS and normal consistency across neighbours is then computed and minimised to obtain the output surface. Further, a new visibility handling approach is proposed for HS datasets. Helmholtz reciprocity constraints are used to compute a confidence score for subsets of cameras, which allows to obtain visibility information.

The proposed approach allows to reconstruct scenes characterised by complex and varying reflectance without prior knowledge or assumptions about it, contrary to established techniques such as MVS and PS. With respect to previous HS formulations, this method does not require a fine initialisation and it computes a solution that is partially guaranteed to be globally optimal. The optimisation is not based on maximum likelihood, which does not allow for surface regularisation and it avoids computing solutions based on local minima which can happen when using iterative approaches. The implementation is also more practical and efficient with respect to the optimisation, contrary to previously proposed approaches where exhaustive search methods were used. Finally, the obtained surface is regularised directly during the MRF optimisation by enforcing consistency across the estimated photometric normals. This offers a strong advantage with respect to using additional surface integration techniques which tend to oversmooth the surface, such as Poisson surface reconstruction. Such techniques do not take full advantage of the normal information obtained through HS due to recomputing the normals for the final topology.

Finally, the results here presented demonstrate that the overall geometric and normal accuracy of this approach significantly outperform previous techniques based on HS. Sub-millimetre accuracy is achieved on all synthetic scenes and normal fidelity is accomplished through the union of surface integration and HS normal estimation. All results were computed on noisy input data, showing that the proposed methodology is robust to it. This is attained as a result of a discriminative regularising term.

In future work, the method could be extended to tackle capture in less controlled scenarios. For example capture in outdoor conditions provides a challenge due to the ambient illumination. Furthermore, use of HS on dynamic scenes has been limited to 2.5D reconstruction [Roubtsova and Guillemaut 2017] to date. The extension to full 3D could be made possible via multi-spectral imaging using a larger number of frequency bands or using temporal multiplexing.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council (Studentship reference number 1815219).

REFERENCES

Gianmarco Addari and Jean-Yves Guillemaut. 2019. An MRF Optimisation Framework for Full 3D Helmholtz Stereopsis. In *14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. 725–736.

Emmanuel A. Audenaert, Jan Van Houcke, Diogo F. Almeida, Lena Paelinck, M. Peiffer, Gunther Steenackers, and Dirk Vandermeulen. 2019. Cascaded statistical shape model based segmentation of the full lower limb in CT. *Computer Methods in Biomechanics and Biomedical Engineering* 22, 6 (2019), 644–657.

Manmohan Chandraker, Jiamin Bai, and Ravi Ramamoorthi. 2013. On Differential Photometric Reconstruction for Unknown, Isotropic BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 12 (Dec 2013), 2941–2955.

Christopher B Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 2016. 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Amaël Delaunoy, Emmanuel Prados, and Peter N. Belhumeur. 2010. Towards Full 3D Helmholtz Stereovision Algorithms. In *Computer Vision - ACCV 2010 - 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12*. 39–52.

Michael Firman, Oisin Mac Aodha, Simon Julier, and Gabriel J. Brostow. 2016. Structured Prediction of Unobserved Voxels From a Single Depth Image. In *CVPR*.

Dan B Goldman, Brian Curless, Aaron Hertzmann, and Steven M. Seitz. 2010. Shape and Spatially-Varying BRDFs from Photometric Stereo. *IEEE Trans. Pattern Anal.*

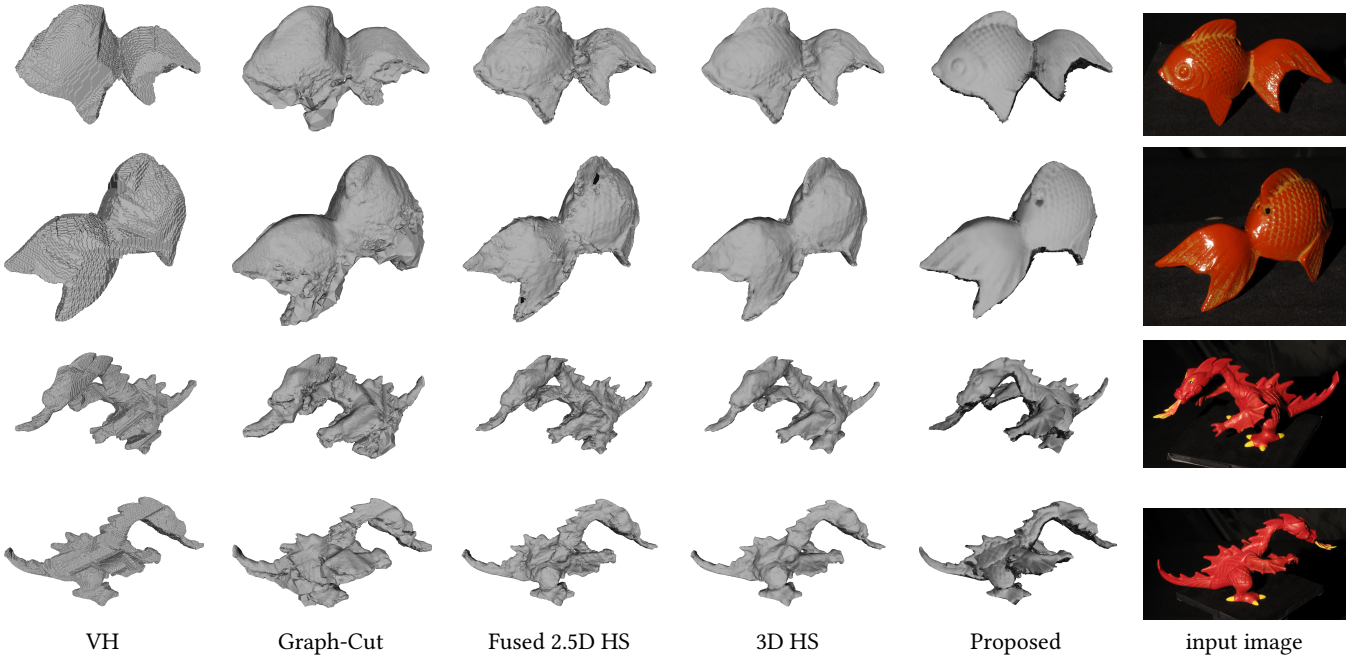


Figure 10: Results obtained on the real scenes from the evaluated methods.

Mach. Intell. 32, 6 (2010), 1060–1071.

Jean-Yves Guillemaut, Ondrej Drbohlav, John Illingworth, and Radim Sára. 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*. 352–359.

Jean-Yves Guillemaut, Ondrej Drbohlav, Radim SÃąra, and John Illingworth. 2004. Helmholtz Stereopsis on Rough and Strongly Textured Surfaces.. In *3DPVT*. IEEE Computer Society, 10–17.

Tian-Qi Han and Hui-Liang Shen. 2015. Photometric Stereo for General BRDFs via Reflection Sparsity Modeling. *IEEE Transactions on Image Processing* 24, 12 (Dec 2015), 4888–4903.

Zsolt Janko, Ondrej Drbohlav, and Radim Sara. 2004. Radiometric calibration of a Helmholtz stereo rig. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1. I–166.

Abhishek Kar, Shubham Tulsiani, João Carreira, and Jagannath Malik. 2015. Category-specific object reconstruction from a single image. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), 1966–1974.

Vladimir Kolmogorov. 2006. Convergent Tree-Reweighted Message Passing for Energy Minimization. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 10 (Oct. 2006), 1568–1583.

Vladimir Kolmogorov. 2013. Reweighted message passing revisited. *CoRR* (2013).

Vladimir Kolmogorov and Carsten Rother. 2007. Minimizing Nonsubmodular Functions with Graph Cuts-A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 7 (July 2007), 1274–1279.

Aldo Laurentini. 1994. The Visual Hull Concept for Silhouette-Based Image Understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* 16, 2 (Feb. 1994), 150–162.

Chen Liang and Kwan-Yee K. Wong. 2010. 3D Reconstruction Using Silhouettes from Unordered Viewpoints. *Image Vision Comput.* 28, 4 (April 2010), 579–589.

Stephen Lombardi and Ko Nishino. 2016. Radiometric Scene Decomposition: Scene Reflectance, Illumination, and Geometry from RGB-D Images. *CoRR* (2016).

Leo Miyashita, Yoshihiro Watanabe, and Masatoshi Ishikawa. 2014. Rapid SVBRDF Measurement by Algebraic Solution Based on Adaptive Illumination. In *2014 2nd International Conference on 3D Vision*, Vol. 1. 232–239.

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

Saulius Nevas, Farshid Manoocheri, and Erkki Ikonen. 2004. Gonioreflectometer for measuring spectral diffuse reflectance. *Appl. Opt.* 43, 35 (Dec 2004), 6391–6399.

Geoffrey Oxholm and Ko Nishino. 2014. Multiview Shape and Reflectance from Natural Illumination. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2163–2170.

Nadejda Roubtsova and Jean-Yves Guillemaut. 2017. Colour Helmholtz Stereopsis for Reconstruction of Dynamic Scenes with Arbitrary Unknown Reflectance. *Int. J. Comput. Vision* 124, 1 (Aug. 2017), 18–48.

Nadejda Roubtsova and Jean-Yves Guillemaut. 2018. Bayesian Helmholtz Stereopsis with Integrability Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 9 (2018), 2265–2272.

Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. 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)*. IEEE Computer Society, Washington, DC, USA, 519–528.

Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. 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 (2008), 1068–1080.

Peter Tu and Paulo R. S. Mendonca. 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.*, Vol. 1. I–541–I–547 vol.1.

George Vogiatzis, Carlos Hernandez, and Roberto Cipolla. 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)*. IEEE Computer Society, Washington, DC, USA, 1847–1854.

George Vogiatzis, Carlos Hernández Esteban, Philip H. S. Torr, and Roberto Cipolla. 2007. Multiview Stereo via Volumetric Graph-Cuts and Occlusion Robust Photo-Consistency. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 12 (Dec. 2007), 2241–2246.

Hermann Von Helmholtz and James PC Southall. 1924. *Helmholtz's treatise on physiological optics, Vol. 1, Trans.* Optical Society of America.

Martin J. Wainwright, Tommi S. Jaakkola, and Alan S. Willsky. 2005. MAP estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory* 51, 11 (Nov 2005), 3697–3717.

Gregory J. Ward. 1992. Measuring and Modeling Anisotropic Reflection. *SIGGRAPH Comput. Graph.* 26, 2 (July 1992), 265–272.

Robert J. Woodham. 1980. Photometric Method For Determining Surface Orientation From Multiple Images. *Optical Engineering* 19, 1 (1980), 191139–191139–.

Todd Zickler. 2006. Reciprocal Image Features for Uncalibrated Helmholtz Stereopsis. In *IEEE Computer Vision and Pattern Recognitiion*. II: 1801–1808.

Todd E. Zickler, Peter N. Belhumeur, and David J. Kriegman. 2002. Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction. *International Journal of Computer Vision* 49, 2 (2002), 215–227.

Todd E. Zickler, Jeffrey Ho, David J. Kriegman, Jean Ponce, and Peter N. Belhumeur. 2003. Binocular Helmholtz stereopsis. In *Proceedings Ninth IEEE International Conference on Computer Vision*. 1411–1417 vol.2.