

# A Bayesian Framework for Enhanced Geometric Reconstruction of Complex Objects by Helmholtz Stereopsis

Nadejda Roubtsova and Jean-Yves Guillemaut

*Centre for Vision, Speech and Signal Processing*

*University of Surrey, Guildford, UK*

*{n.roubtsova, j.guillemaut}@surrey.ac.uk*

**Keywords:** 3D reconstruction, Helmholtz stereopsis, complex reflectance

**Abstract:** Helmholtz stereopsis is an advanced 3D reconstruction technique for objects with arbitrary reflectance properties that uniquely characterises surface points by both depth and normal. Traditionally, in Helmholtz stereopsis consistency of depth and normal estimates is assumed rather than explicitly enforced. Furthermore, conventional Helmholtz stereopsis performs *maximum likelihood* depth estimation without neighbourhood consideration. In this paper, we demonstrate that reconstruction accuracy of Helmholtz stereopsis can be greatly enhanced by formulating depth estimation as a Bayesian *maximum a posteriori probability* problem. In reformulating the problem we introduce neighbourhood support by formulating and comparing three priors: a depth-based, a normal-based and a novel depth-normal consistency enforcing one. Relative performance evaluation of the three priors against standard maximum likelihood Helmholtz stereopsis is performed on both real and synthetic data to facilitate both qualitative and quantitative assessment of reconstruction accuracy. Observed superior performance of our depth-normal consistency prior indicates a previously unexplored advantage in joint optimisation of depth and normal estimates.

## 1 INTRODUCTION

As evidenced by the formidable volume of past and active research, reconstruction of 3D geometry is both challenging and much desirable for practical applications. A tremendous progress has been made in the field with sub-millimetre accurate geometries being obtained when capture conditions and surface properties are tailored for reconstruction. All prior algorithms rely on multiple images to resolve inherent depth ambiguity, while differing in acquisition and view correlation employed to formulate ambiguity resolving constraints. Also variable is the degree of neighbour support used when characterising a surface point by its depth, normal or both.

The oldest of 3D reconstruction techniques is conventional stereopsis thoroughly surveyed in (Scharstein and Szeliski, 2002) and (Seitz et al., 2006) for single- and multi-view respectively. Conventional stereopsis computes disparity through feature-point intensity matching between acquired views in the presence of sufficient texture. In this approach, surface points are characterised by depth, which is reciprocal to disparity. Conventional intensity-based stereo strongly relies on the inter-viewpoint constancy

of feature appearance (i.e. Lambertian reflectance) failing when the assumption is violated. Alternative SIFT features are robust to intensity variations, although they only facilitate sparse representation.

Another reconstruction approach called photometric stereopsis (Woodham, 1989) permits an arbitrary reflectance model as long as it is *a priori* known. Photometric constraints are linked to the response of a point to varying illumination at a *constant* viewpoint. The sought surface orientation is the one best reconciling intensity predictions and measurements. Essentially, the reflectance issue of conventional stereo is not solved by photometric stereo but rather the burden of it is shifted to the calibration phase of which a surface-orientation-dependent reflectance model is required. Reflectance modelling is a tedious task, often impossible to the desired accuracy for real objects. With an accurate model, photometric stereo directly outputs highly descriptive surface normals. Individual normals however need to be integrated into a surface often resulting in drift (global shape distortion) due to accumulation of numerical integration errors. Unlike conventional stereo, photometric stereo with its single viewpoint avoids the task of feature-point matching.

There are few techniques bypassing the need for

reflectance modelling. One old example is the shape-from-silhouette algorithm (Baumgard, 1974) which computes a rough 3D outline of the object, its visual hull (Laurentini, 1994), by intersecting visual cones of multiple views. Visual hull is used for initialisation by many advanced reconstruction algorithms. A more recent highly promising photometric technique of Helmholtz stereopsis (HS) (Zickler et al., 2002) addresses the fundamental problem of reflectance modelling by enforcing consistency of *reflectance-model-independent* Helmholtz reciprocity observation: i.e. swapping the sensor and the light source does not alter reflectance response. Besides its independence of the reflectance model, HS has the unique feature of point characterisation by both depth and normal. However, in its conventional formulation, HS is sequential with depth estimates uniquely determining the normals. Such unidirectional indexation of normals by depth estimates means that the typically noisy HS depth maps result in normal inaccuracies and hence in local and global reconstruction errors (although a certain degree of robustness to normal errors has been observed). Conceptually, the depth-normal dependency need not be unidirectional: depth and normal estimation can be unified in a single framework enforcing consistency and comparable accuracy levels of both estimates.

Conventional HS essentially performs *maximum likelihood estimation* (MLE). Even in the absence of noise, inherent point depth ambiguities exist, for instance, due to coincidental symmetries in the acquisition set-up configuration relative to the sampled surface. Since local evidence is ambiguous, neighbourhood support is clearly needed and warranted as there is always a degree of local smoothness in real objects. Yet, to our knowledge, MAP optimisation in the context of HS has not been previously attempted. In this paper, we propose a novel MAP formulation embedding HS into a Bayesian framework with a prior that for the first time explicitly enforces consistency between depth and normals. We show that, with the consistency prior, MRF optimisation of MAP HS indeed results in superior reconstruction accuracies. Unlike alternative depth-based or normal-based priors, the depth-normal consistency prior capitalises on the unique ability of HS to provide both depths and normals and, combining the two, produces the most correct geometries coherent in both depth and integrated normal representation. These conclusions are based on the quantitative and qualitative evaluation comparing conventional ML HS to MAP optimisation with 1. classical depth-based; 2. normal-based and 3. novel depth-normal consistency priors.

## 2 RELATED WORK

Helmholtz reciprocity states that a light ray and its reverse will undergo the same processes of reflection, refraction and absorption (Helmholtz, 1925). Let  $\hat{v}_1$  be the unit vector directed from the surface point to the camera and  $\hat{v}_2$  the corresponding vector from the surface point to the light source. The implication of Helmholtz reciprocity, first observed by Zickler *et al.* (Zickler et al., 2002) in the context of multi-view reconstruction, is that interchanging the light source and camera in the set-up, thereby swapping the vector definitions, has no effect on the point's reflective behaviour. Mathematically, Bidirectional Reflectance Distribution Function (BRDF)  $f_r$  is reciprocal:  $f_r(\hat{v}_2, \hat{v}_1) = f_r(\hat{v}_1, \hat{v}_2)$ . The following standard image formation equations for reciprocal images  $I_1$  and  $I_2$  respectively:

$$i_1 = f_r(\hat{v}_2, \hat{v}_1) \frac{\hat{n} \cdot \hat{v}_2}{r_2}, \quad i_2 = f_r(\hat{v}_1, \hat{v}_2) \frac{\hat{n} \cdot \hat{v}_1}{r_1} \quad (1)$$

express a surface point's image intensities  $i_1$  and  $i_2$  as a function of BRDF, surface normal  $\hat{n}$ , the two reciprocal unit vectors and the radiation fall-off factor  $r$ . Reciprocity of BRDF in conjunction with Equation 1 result in the following constraint  $\hat{w}$  notably *without* any dependency on the BRDF:

$$\left( i_1 \frac{\hat{v}_1}{r_1} - i_2 \frac{\hat{v}_2}{r_2} \right) \cdot \hat{n} = \hat{w} \cdot \hat{n} = 0. \quad (2)$$

With a single  $\hat{w}$  per reciprocal pair, 3 or more reciprocal pairs result in constraint matrix  $W$  to which singular value decomposition (SVD) can be applied:  $SVD(W) = U\Sigma V^*$  where  $U, V$  are unitary and  $\Sigma$  is a rectangular diagonal matrix. The last column of  $V$  gives the normal at the sampled point. The last diagonal value of  $\Sigma$ , the SVD residual  $\sigma_3$ , tends to 0 when there is a mutual constraint consistency. For outlier elimination, Zickler *et al.* also involve  $\sigma_2$  in their consistency measure: the SVD coefficient  $\frac{\sigma_2}{\sigma_3}$  tends to infinity for consistent  $W$  (those of true surface points).

The requirement of multiple reciprocal pairs and the need for careful offline calibration led to discussions on acquisition impracticality of HS. In response, Zickler in (Zickler, 2006), firstly, devises an *auto-calibration* algorithm using specular highlights and intensity patches, two inherent easily identifiable regions of interest. Unlike conventional stereo, intensity matching in HS is not conditional on the validity of the Lambertian assumption making intensity patches stable calibration markers.

The inherent appearance predictability of an intensity patch of one reciprocal image based on the other is also employed in (Tu et al., 2003) to formulate prediction error for registration in full 3D HS.

Work on full 3D HS is scarce. The work of Delaunoy *et al.* (Delaunoy *et al.*, 2010) and the more recent publication of Weinmann *et al.* (Weinmann *et al.*, 2012) are two notable examples. Both are variational approaches requiring computationally intensive optimisation over the entire surface with long execution times. The method of Weinmann *et al.* is more cumbersome due to fusion with structured light at acquisition. The impracticalities are however outweighed by the impressive degree of demonstrated reconstruction detail.

To address the issue of multiple constraint requirement, Zickler *et al.* in (Zickler *et al.*, 2003) propose binocular HS i.e reconstruction from a single reciprocal pair. The method is a differential approach where the (single) constraint is formulated as a PDE of depth over the surface coordinates. The PDE requires initialisation and results in a family of solutions the ambiguity of which is resolved through multi-pass optimisation. Although interesting as an exercise on acquisition simplification, additional computational complexity is perhaps not outweighed by the advantages.

Binocular HS and full 3D HS methods exploit the advantages of optimisation over a set of surface points. In the original HS, the depth label at each point is computed independently of its neighbours (without a Bayesian *prior*). Logically, there is a strong correlation between neighbouring surface points. By formulating a prior, the problem is turned into a MAP one, solvable by numerous mature MRF optimisation techniques (Szeliski *et al.*, 2008). The value of MRF optimisation for other 3D reconstruction approaches has been established. In conventional stereo, the top-performers are global optimisation algorithms solved through MRF-based graph-cuts and belief propagation (Scharstein and Szeliski, 2002). (Wu *et al.*, 2006) is an interesting work where MRF optimisation is performed in the context of photometric stereo with a normal-based (rather than depth-based) prior achieving remarkable robustness in the face of noisy input, complex geometries, shadows, transparencies etc..

Conventional HS will similarly benefit from MRF-based optimisation, since geometries obtained by ML HS lack in smoothness and fine structural detail due to inherent depth ambiguities, intensity measurement noise, sensor saturation and calibration/discretisation errors. Aimed at tackling the issues of conventional HS, the contributions of this paper are twofold. Firstly, we devise a MAP framework allowing to apply MRF-based optimisation in the context of HS. Secondly, we introduce a novel smoothness prior enforcing coherence of depth and normal estimates of

HS and show superiority of the prior over the purely depth-based or normal-based ones.

### 3 METHODOLOGY

In reconstruction by HS, we begin with  $N$  reciprocal image pairs ( $N \geq 3$ ) and a discrete volume  $V$  of  $N_X \times N_Y \times N_Z$  voxels  $v(x, y, z)$  containing the object. Each  $v(x, y, z)$  is sampled by projection onto the reciprocal images to acquire a set of  $N$  intensity 2-tuples  $\{(i_1, i_2)_1, \dots, (i_1, i_2)_N\}$  and formulate  $N$  constraints  $\hat{w}$  as in Equation 2. Only those  $v(x, y, z) \in V$  containing surface points will have  $N$  mutually consistent constraints. In standard HS, voxel sets  $\mathcal{R}_{p(x,y)} = \{v(x, y, z) : x = x^*, y = y^*\}$  with constant 2D coordinates  $x^*$  and  $y^*$  are sampled exhaustively over  $V$ . Each  $\mathcal{R}_{p(x,y)}$  defines the depth search space for random variable  $p(x, y)$  with the optimal depth value  $z_{p(x,y)} = z_{p(x,y)}^*$  corresponding to the surface point  $P(x, y, z_{p(x,y)}^*)$ . In our work, the object's visual hull (VH) restricts the search space to  $\mathcal{R}_{p(x,y)}^{VH} = \{v(x, y, z) : x = x^*, y = y^*, v(x^*, y^*, z_{p(x,y)}) \in VH\}$  (implicit in Equations 4 and 5) limiting sampling in both 2D ( $\mathcal{R}_{p(x,y)} \cap VH = \emptyset$  for some  $p(x, y)$ ) and depth.

We postulate that depth estimation of HS is in fact a *labelling* problem where each random variable  $p(x, y)$  is assigned depth label  $z_{p(x,y)}^*$ . The optimal solution to such a labelling problem is known (Li, 1994) to be one maximising the *a posteriori* probability (MAP) i.e. the likelihood of a parameter assuming a certain value given the observation. The problem is typically translated into the equivalent one of posterior energy minimisation where the total energy is a sum of likelihood (data) and prior energy terms. Likelihood is related to the noise model of the observation representing its quality, viewed independently of the other observations. The prior term (typically local smoothness) is the knowledge of the problem encapsulating interaction between observations. In our work, the interaction is contained within the Markovian neighbourhood of  $p(x, y)$  restricted by VH:

$$\mathcal{N}(p(x, y)) = \{p(x+k, y+(1-k)) : k \in \{0, 1\}, \mathcal{R}_{p(x+k, y+(1-k))}^{VH} \neq \emptyset\} \quad (3)$$

In this paper, we formulate depth estimation of HS as a MAP problem in Equation 4. For each random variable  $p(x, y)$  and given  $p(x', y') \in \mathcal{N}(p(x, y))$ , we define normalised data and smoothness costs,  $E_d(x, y, z_{p(x,y)})$  and  $E_s(x, y, z_{p(x,y)}, x', y', z_{p(x', y')})$  respectively, weighted by the normalised parameter  $\alpha$ . The total optimisation is over all random variables within VH. The solution to the labelling problem is

the label configuration  $f^* = \{z_{p(x,y)}^* \in [Z_1, \dots, Z_{N_Z}] : x \in [X_1, \dots, X_{N_X}], y \in [Y_1, \dots, Y_{N_Y}]\}$  selected from a set of  $\mathcal{S}$  such configurations.

$$f_{MAP}^* = \arg \min_{f \in \mathcal{S}} \sum_{p(x,y)} ((1 - \alpha) E_d(x, y, z_{p(x,y)}) + \alpha E_s(x, y, z_{p(x,y)}, x', y', z_{p(x',y')})) \quad (4)$$

In contrast to our approach, conventional HS solves a simpler *maximum likelihood* (ML) optimisation problem, without the smoothness prior, resulting a sub-optimal solution because each random variable is optimised independently:

$$f_{ML}^* = \arg \min_{f \in \mathcal{S}} \sum_{p(x,y)} E_d(x, y, z_{p(x,y)}) \quad (5)$$

Sub-optimality leads to noisy depth maps and hence lacking surface smoothness and structural detail. The global shape may be reasonable, but the reconstruction finesse of conventional HS is limited because noisy depth labels index approximate normals. Through Bayesian formulation in Equation 4, we endeavour to obtain cleaner depth maps improving accuracy by more accurate normal indexing. A Bayesian framework is clearly more suitable because of the strong statistical dependency between neighbouring depth estimates. The following sections define data energy and three investigated smoothness priors.

### 3.1 Data term

Depth hypothesis  $z$  of each  $p(x, y)$  has got a likelihood  $E_d(x, y, z)$  defined through the SVD coefficient  $\frac{\sigma_2(v(x, y, z))}{\sigma_3(v(x, y, z))}$  associated with the corresponding voxel. The coefficient tends to infinity as  $z$  approaches the correct depth  $z^*$ . Since MRF optimisation is formulated as a minimisation,  $E_d(x, y, z)$  is a decaying function of  $\frac{\sigma_2(v(x, y, z))}{\sigma_3(v(x, y, z))}$ . Throughout the paper, we adhere to the exponential formulation with the decay factor  $\mu = 0.2 \times \log(2)$ :

$$E_d(x, y, z) = e^{-\mu \times \frac{\sigma_2(v(x, y, z))}{\sigma_3(v(x, y, z))}} \quad (6)$$

As it indicates likelihood, the function is bounded in the range  $[0, 1]$ .

### 3.2 Smoothness term

Unlike conventional *data-based* HS, we introduce priors. We devise three prior types reflecting the unique ability of HS to generate both depth and normal estimates.

**1. Depth-based prior.** The prior is known from conventional stereo. In this work, we define depth-based smoothness cost  $E_{s,d}(x, y, z, x', y', z')$  of voxel  $v(x, y, z)$  relative to  $v(x', y', z')$  by the discontinuity-preserving truncated squared difference of the neighbouring depth hypotheses normalised by the total number of labels ( $N_Z$ ):

$$E_{s,d}(x, y, z, x', y', z') = \lambda \times \min(S_{max}, (z - z')^2) \quad (7)$$

where  $S_{max} = (0.5 \times N_Z)^2$  is the truncation value and  $\lambda = (N_Z)^{-2}$  is the normalising constant. With penalties for much different labels at neighbouring surface points, the prior encourages piece-wise constant depth and biases towards a fronto-parallel representation. Discontinuities are preserved by truncation moderating depth fluctuation penalties.

**2. Normal-based prior.** Surface characterisation through normals is typical of photometric techniques. A suitable normal-based prior would enforce locally constant normals encouraging locally flat, though not necessarily fronto-parallel, surfaces. Hence, this prior is less restrictive of reconstructed surfaces than the depth-based one. The complication in formulating normal-based priors arises because 1. normals are continuous 3D quantities that cannot be labels in our discrete framework and 2. normal correlations are irregular expressions not optimisable by graph cuts (Kolmogorov and Zabih, 2004). In this work, discrete depths are still the labels, but rather than depth information, we use the corresponding normal similarity to assess label compatibility. Sequential tree re-weighted message passing (TRW-S) (Wainwright et al., 2005), (Kolmogorov, 2006) MRF optimisation is used consistently throughout the paper because it does not require regularity of prior.

Given photometric normal  $\hat{n}(x, y, z) = (n_x, n_y, n_z)^\top$  and  $\hat{n}(x', y', z') = (n'_x, n'_y, n'_z)^\top$  of voxels  $v(x, y, z)$  and  $v(x', y', z')$  respectively, we formulate the smoothness-based constraint:

$$E_{s,n}(x, y, z, x', y', z') = \pi^{-1} \arccos(\hat{n}(x, y, z) \cdot \hat{n}(x', y', z')) \quad (8)$$

The cost function of Equation 8 is the normalised correlation angle between normals.

**3. Depth-normal consistency prior.** The depth-based prior seeks to clean up depth maps by enforcing their smoothness, while the normal-based approach promotes gradual spatial evolution of the normal field. Both approaches are one-sided: the depth is optimised indexing the normals or vice versa. Depth and normal estimation processes are however not independent and must be consistent with each other. We formulate a prior explicitly enforcing consistency between depths and normals, for the first time performing joint depth map and normal field optimisation.



Figure 1: Real data: object appearance.

Each depth transition between  $v(x, y, z)$  and  $v(x', y', z')$  uniquely defines local *geometric* normal  $\hat{n}_g(x, y, z, x', y', z')$ , always contained in the transition plane ( $xz$  or  $yz$ ). If the depth transition is correct, the geometric normal correlates well with the projections of the photometric normals of  $v(x, y, z)$  and  $v(x', y', z')$  (their normal estimates), respectively  $\hat{n}_{ph-prj}(x, y, z)$  and  $\hat{n}_{ph-prj}(x', y', z')$ , onto the corresponding depth transition plane. Mathematically, the correlation degree can be expressed by the correlation angle. For example, for  $v(x, y, z)$  we have:

$$\phi_{ph-g} = \pi^{-1} \arccos(\hat{n}_{ph-prj}(x, y, z) \cdot \hat{n}_g(x, y, z, x', y', z')) \quad (9)$$

Since the range of the arc-cosine function is  $[0, \pi]$ , the orientation of the geometric normal must be forced to be consistent with the photometric normals (i.e. out of the surface,  $z > 0$ ). Hence, the depth-normal consistency prior  $E_{s-dn}(x, y, z, x', y', z')$  is formulated as follows:

$$E_{s-dn}(x, y, z, x', y', z') = \frac{1}{2}(\phi_{ph-g} + \phi'_{ph-g}) \quad (10)$$

## 4 EVALUATION

We perform evaluation of the proposed method on both real and synthetic data to enable qualitative and quantitative analysis. Throughout, our method with its 3 prior options is compared against the standard HS approach without MRF optimisation. For real data, the quality of results is assessed visually as there is no ground truth. Synthetic input imagery, on the other hand, is generated from an a priori known mesh and hence permits quantitative assessment of both local and global shape deviations of the reconstructions.

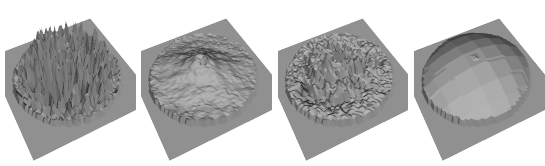
### 4.1 Real data

Real data is composed of 4 sets (Figure 1) from (Guillemaut et al., 2008), each posing different challenges. The billiard ball and the teapots are specular smooth objects. The teapots are more complex with wider specularities and 2D texture (e.g. stripes, flowers). The terracotta doll is Lambertian but has many fine geometric details (e.g. dimples, clothing).

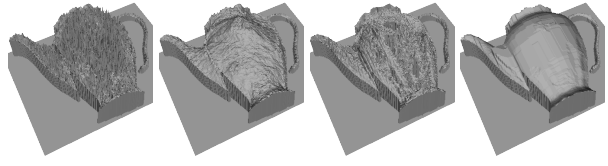
Figure 2 contrasts estimated depth maps and the corresponding integrated surfaces of the proposed MAP HS formulation against standard (ML) HS. MAP optimisation priors are compared by qualitatively accessing both local and global accuracy of the generated depth maps and surfaces. The relative weight  $\alpha$  is tuned for each prior independently but the optimal setting per prior tends to be consistent across all datasets. Surface integration is performed from the normal fields in the frequency domain using the Frankot-Chellappa (FC) algorithm (Frankot and Chellappa, 1988).

The results in Figure 2 show that, relative to ML HS, MAP optimisation clearly produces smoother reconstructions (teapot/ball surface in Figures 2(e),(f),(g)) with finer structural detail (the doll's dimples/eyebrows in Figure 2(h); the corresponding normal fields in Figure 2(m) are rectified to reveal structural detail). On the relative performance of the priors, the key observation is that only the depth-normal consistency prior generates geometrically correct depth maps. Global accuracy of the depth-based prior ranges from poor (ball, Figure 2(a)) to reasonable (teapots, Figures 2(b),(d)) and high (doll, Figure 2(c)). The depth maps of the depth-based prior are however universally noisy. Corresponding normal-based prior results are consistently heavily distorted. Correct reconstruction is only possible from a geometrically (globally and locally) accurate depth map. While this is evident for the doll dataset, in other cases distorted depth maps of the normal-based approach may, in the front view, seem to produce decent reconstructions (Figures 2(f),(g)). The deceiving appearance results from optimisation accidentally finding well-correlating normals at wrong depths. These integrate into locally smooth surfaces, yet distort the global shape for all four datasets (Figures 2(i),(j),(k),(l)). As the depth map accuracy assessment suggests, the best global shape reconstruction belongs to the depth-normal consistency prior optimisation.

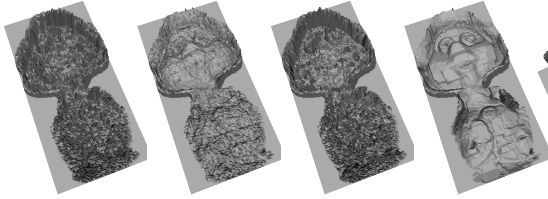
HS is known (Guillemaut et al., 2004) to be sensitive to calibration and discretisation errors whereby, due to projection mismatches, intensity measurements  $i_1$  and  $i_2$  within a reciprocal pair are inconsistent. The grave form of inconsistency when, within a single reciprocal pair, a point is projected into different intensity fields, results in reconstruction of 2D (imprinted) texture as geometric detail. Along with sharpening true geometries, optimisation is seen to strengthen the effect of calibration errors: e.g. printed stripes and flowers on the teapots (Figure 1) appear embedded into geometry (Figure 2(f),(g)). Performing patch-based intensity averaging during input data



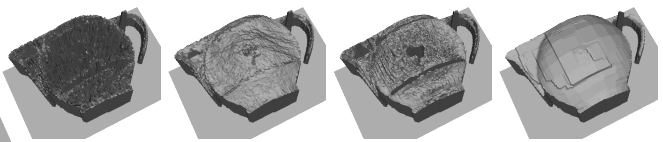
(a) depth maps,  $\alpha_d = 0.9, \alpha_n = 0.1, \alpha_{dn} = 0.8$



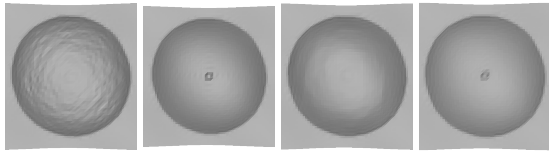
(b) depth maps,  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$



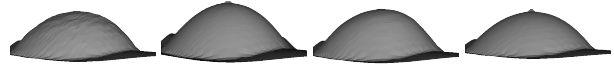
(c) depth maps,  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$



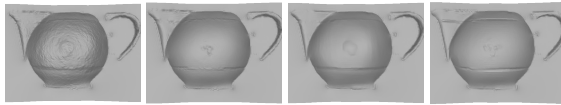
(d) depth maps,  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$



(e) FC reconstruction, front view,  $\alpha_d = 0.9, \alpha_n = 0.1, \alpha_{dn} = 0.8$



(i) FC reconstruction, side view,  $\alpha_d = 0.9, \alpha_n = 0.1, \alpha_{dn} = 0.8$



(f) FC reconstruction, front view,  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$



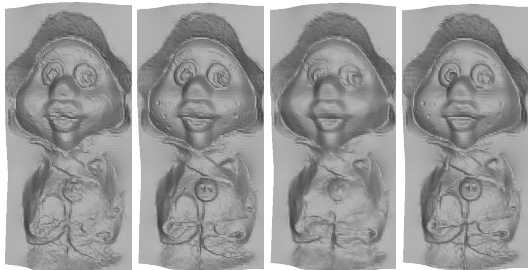
(j) FC reconstruction, bottom view,  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$



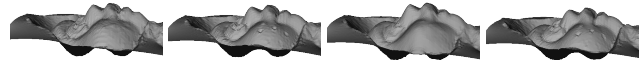
(g) FC reconstruction, front view,  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$



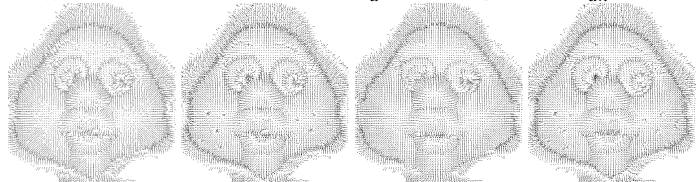
(k) FC reconstruction, bottom view,  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$



(h) FC reconstruction, front view,  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$



(l) FC reconstruction, side view,  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$



(m) Normal fields (doll),  $\alpha_d = 0.9, \alpha_n = 0.5, \alpha_{dn} = 0.8$

Figure 2: Depth maps, normal fields (doll) and final reconstruction by integration using the FFT-based Frankot-Chellappa (FC) algorithm. In each sequence of 4 images (left to right) standard (ML) HS is compared against proposed MAP HS formulation using MRF optimisation with depth-based ( $d$ ), normal-based ( $n$ ) and depth-normal consistency ( $dn$ ) priors (data-smoothness weighting  $\alpha$  as indicated in each case). Sampling resolution  $\Delta x \times \Delta y \times \Delta z$  and sampled volumes  $|V| = N_x \times N_y \times N_z$  are as follows. Doll:  $1.0mm \times 1.0mm \times 0.5mm$ ,  $|V| = 160 \times 82 \times 60$ ; teapot no.1:  $1.0mm \times 1.0mm \times 0.25mm$ ,  $|V| = 150 \times 200 \times 320$ ; teapot no.2:  $1.0mm \times 1.0mm \times 0.25mm$ ,  $|V| = 120 \times 190 \times 480$ ; billiard:  $1.0mm \times 1.0mm \times 0.25mm$ ,  $|V| = 60 \times 60 \times 100$ .

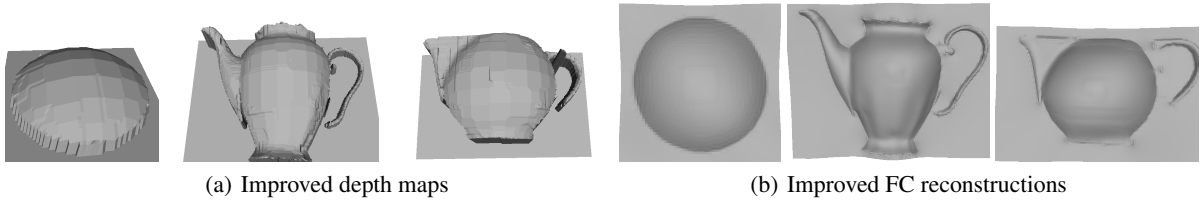


Figure 3: Mitigation of saturation and mis-calibration artefacts through patch-based averaging. Settings: MAP HS with depth-normal consistency prior, patch-based averaging and  $\alpha_{dn} = 0.5$ .

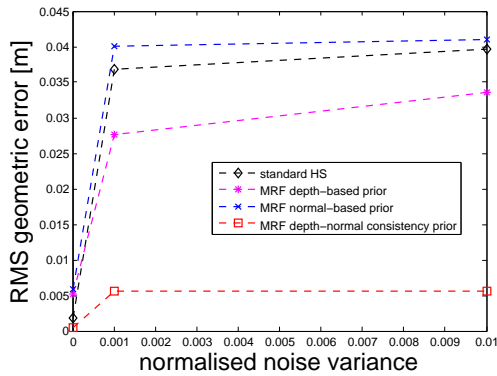


Figure 4: Quantitative evaluation on synthetic data. RMS geometric error as a function of normalised intensity noise variance for our MRF (MAP) HS formulation with different priors and standard (ML) HS.

sampling to bring inconsistent measurements into closer correspondence would counteract the problem (Guillemaut et al., 2004). The measure mitigates saturation artefact as well since, instead of the indiscriminative sensor upper limit, the intensities in Equation 2 acquire more meaningful values through regional support of patch-based averaging. In Figure 3(a) we show how patch-based averaging effectively eliminates depth errors in the saturated and pattern regions, hence eliminating/substantially mitigating saturation and embedded 2D texture artefacts (Figure 3(b)).

## 4.2 Synthetic data

Our synthetic object of choice is a sphere (radius = 200 mm). The mesh of the sphere is rendered in POV-Ray (POV-Ray, 2013) with reflectance consisting of diffuse and specular components and under controlled camera/light source configurations to produce 8 noise-free reciprocal pairs. For robustness assessment, two additional image sets are produced by corrupting the original set by Gaussian zero-mean *intensity* noise with the variance levels of respectively 0.001 and 0.01. Reconstruction quality is quantified

by the geometric error: the reconstructed mesh is sampled for each point seeking the closest distance on the target ground truth mesh. Figure 4 shows the RMS error as a function of noise variance comparing standard ML HS and our MRF (MAP) formulation with 3 different priors (sampled volume  $|V| = 82 \times 82 \times 251$  at resolution  $5mm \times 5mm \times 1mm$ ). The input intensity range is  $[0, 65535]$ . Hence, the normalised noise variance of 0.001 and 0.01 translates into the absolute intensity standard deviations of 2072 and 6554 respectively. Surface integration is performed using unscreened Poisson surface reconstruction (Kazhdan et al., 2006) because the method outputs surfaces characterised in the absolute world coordinate system facilitating their easy comparison unlike FC surfaces in their individual relative reference frames needing alignment.

In Figure 4 MRF optimisation with the depth-normal consistency prior clearly outperforms the other approaches by an order of magnitude margin achieving reconstruction accuracies of roughly 0.5mm and 5mm for noise-free and noise-corrupted sets respectively. The accuracy of the other methods is the order of a few mm and a few cm for the two respective cases. For the noise-free case, standard HS comes closest to the depth-normal prior with the 2mm error. The second observation to be made is that the normal-based prior is utterly ineffective with the accuracy always below standard HS, consistent global shape distortion and the greatest performance deterioration with increased noise levels owing to the high susceptibility of normals as continuous 3D quantities to input noise via SVD. Even for the noise-free case, the error with the normal-based prior is high because there is no theoretical guarantee that the best correlating normals correspond to the correct depth positions. The depth-based prior appears to facilitate an improvement compared to standard HS for noise-corrupted data. On visual inspection however, the depth-based reconstruction is seen to retain the global shape distortion of the standard HS result, albeit mitigated as its lower RMS indicates. In the noisy case, only our depth-normal consistency prior produces a

global shape visually acceptable as a sphere.

## 5 CONCLUSIONS

We have proposed a novel MRF optimisable MAP formulation of HS, instead of its standard ML form. To this end, we formulated and compared a depth-based, a normal-based and a specially tailored depth-normal consistency prior. We conclude that correctly utilising the given of piece-wise surface smoothness in the MAP formulation greatly improves both local and global reconstruction accuracy relative to ML results. Both quantitative and qualitative results indicate our depth-normal consistency prior to be the correct formulation of the smoothness term, which by enforcing consistency between depth and normal information produces the best results in terms of both local smoothness and global object shape. The results generated with the prior are uniquely consistent in both depth and integrated normal domain with the normals being indexed from a geometrically correct depth map. The computational overhead for the prior is dependent on the size of the sampled voxel volume and for the presented real data ranges from 2 minutes (billiard ball) to 4 hours (teapot no. 2). In our future work we are confident the run-times will be reduced substantially by embedding MRF optimisation of Bayesian HS into a coarse-to-fine framework using octrees and/or by parallelising and porting prior cost pre-computation (the bottleneck of the pipeline) onto the GPU. In addition, we shall explore the potential of our MRF framework in resolving ambiguities in the sensor saturation region and at grazing sampling angles.

## REFERENCES

- Baumgard, B. (1974). *Geometric Modeling for Computer Vision*. PhD thesis, University of Stanford.
- Delaunoy, A., Prados, E., and Belhumeur, P. (2010). Towards full 3D Helmholtz stereovision algorithms. In *Proc. of ACCV*, volume 1, pages 39–52.
- Frankot, R. and Chellappa, R. (1988). A method for enforcing integrability in shape from shading algorithms. *PAMI*, 10(4):439–451.
- Guillemaut, J.-Y., Drbohlav, O., Illingworth, J., and Šára, R. (2008). A maximum likelihood surface normal estimation algorithm for Helmholtz stereopsis. In *Proc. of VISAPP*, volume 2, pages 352–359.
- Guillemaut, J.-Y., Drbohlav, O., Šára, R., and Illingworth, J. (2004). Helmholtz stereopsis on rough and strongly textured surfaces. In *Proc. of 3DPVT*, pages 10–17.
- Helmholtz, H. (1925). *Treatise on Physiological Optics*, volume 1. Dover (New York).
- Kazhdan, M., Bolitho, M., and Hoppe, H. (2006). Poisson surface reconstruction. In *Proc. of SGP*, pages 61–70.
- Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568 – 1583.
- Kolmogorov, V. and Zabih, R. (2004). What energy functions can be minimized via graph cuts? *PAMI*, 26:65–81.
- Laurentini, A. (1994). The visual hull concept for silhouette-based image understanding. *PAMI*, 16(2):150–162.
- Li, S. (1994). Markov random field models in computer vision. In *Proc. of ECCV*, volume B, pages 361–370.
- POV-Ray (2013). POV-Ray - The Persistence of Vision Raytracer. <http://www.povray.org/>.
- Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1-3):7–42.
- Seitz, S., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. of CVPR*, volume 1, pages 519– 528.
- 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. *PAMI*, 30(6):1068–1080.
- Tu, P., Mendonça, P. R., Ross, J., and Miller, J. (2003). Surface registration with a helmholtz reciprocity image pair. In *Proc. of IEEE Workshop on Color and Photometric Methods in Computer Vision*.
- Wainwright, M. J., Jaakkola, T. S., and Willsky, A. S. (2005). Map estimation via agreement on trees: Message-passing and linear-programming approaches. *IEEE Transactions on Information Theory*, 51(11):3697–3717.
- Weinmann, M., Ruiters, R., Osep, A., Schwartz, C., and Klein, R. (2012). Fusing structured light consistency and helmholtz normals for 3D reconstruction. In *Proc. of BMVC*, pages 108.1–108.12. BMVA Press.
- Woodham, R. J. (1989). *Shape from shading*, chapter Photometric method for determining surface orientation from multiple images, pages 513–531. MIT Press, Cambridge, MA, USA.
- Wu, T.-P., Tang, K.-L., Tang, C.-K., and Wong, T.-T. (2006). Dense photometric stereo: A Markov Random Field approach. *PAMI*, 28(11):1830–1846.
- Zickler, T. (2006). Reciprocal image features for uncalibrated Helmholtz stereopsis. In *Proc. of CVPR*, pages 1801– 1808.
- Zickler, T., Belhumeur, P. N., and Kriegman, D. J. (2002). Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *IJCV*, 49(2–3):215–227.
- Zickler, T. E., Ho, J., Kriegman, D. J., Ponce, J., and Belhumeur, P. N. (2003). Binocular Helmholtz stereopsis. In *Proc. of ICCV*, volume 2, pages 1411–1417.