3D Reconstruction in Laparoscopy with Close-Range Photometric Stereo

Toby Collins and Adrien Bartoli

ALCoV-ISIT
CNRS and Université d'Auvergne, Clermont-Ferrand, France
Toby.Collins@gmail.com, Adrien.Bartoli@gmail.com

Abstract. In this paper we present the first solution to 3D reconstruction in monocular laparoscopy using methods based on Photometric Stereo (PS). Our main contributions are to provide the new theory and practical solutions to successfully apply PS in close-range imaging conditions. We are specifically motivated by a solution with minimal hardware modification to existing laparoscopes. In fact the only physical modification we make is to adjust the colour of the laparoscope's illumination via three filters placed at its tip. Once calibrated, our approach can compute 3D from a single image, does not require correspondence estimation, and computes absolute depth densely. We demonstrate the potential of our approach with ground truth ex-vivo and in-vivo experimentation.

1 Introduction

An important computer vision task in Minimally Invasive Surgery (MIS) is to recover the 3D structure of organs and tissues viewed in endoscopic images and videos. A general solution to this has many important applications, including enhanced intra-operative surgical guidance, depth perception, 3D motion estimation and compensation, novel-view synthesis and improving pre-operative/intra-operative data registration. In the literature, the main practical monocular reconstruction approaches so far are based on Structure-from-Motion (SfM). However, since this is correspondence based, it is error prone and at textureless regions 3D cannot be recovered. SfM also requires very strong assumptions on surface motion (e.g. rigid or periodic motion), and requires sufficient motion baseline. By contrast, PS offers a very different approach for 3D which is based on photometric constraints using three or more light sources [2, 15, 8]. PS is attractive since it provides dense 3D estimates, does not require correspondence estimation, and can compute 3D from just a single colour image. However, to date PS has not been applied to 3D reconstruction in MIS. Our main contributions are to provide the theory and practical solutions to successfully apply PS to the very close-range imaging conditions of MIS. In this paper we focus on laparoscopy. On the hardware side, our solution takes a standard monocular laparoscope, modified only with three colour filters (red, green and blue) placed at its tip. This corresponds to a practical and very inexpensive modification. The physical dimensions remain unchanged and it does not require any strobing or synchronised triggering between the camera and light source.

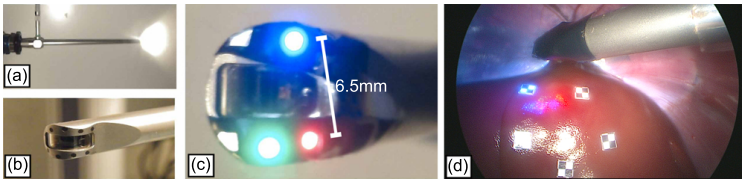

Fig. 1: Modification made to a standard laparoscope (a,b) to facilitate practical in-vivo photometric stereo by fitting three colour filters at its tip (c). Photograph of in-vivo tests (d).

3D Reconstruction in MIS. Several different research directions for 3D reconstruction in MIS have emerged over the recent years. These differ in the sensing hardware required to compute 3D. At one end of the spectrum are dedicated 3D sensing devices. These have included Structured Light (SL) setups [1] and Time-of-Flight (ToF) cameras [11]. SL requires additional instruments, which may clutter the scene and to date, neither SL nor ToF sensors have proved sufficiently reliable in practice. Stereo endoscopes have also been tried for 3D reconstruction [13, 4, 7]. While promising, these are limited by fixed camera convergence angles and stereo baseline, and perform poorly at textureless regions. At the other end of the spectrum are passive monocular methods. These require no additional instruments and compute 3D using the raw video feed. This however is an extremely challenging computer vision problem. Some progress has been made using SfM and Simultaneous Localisation And Mapping (SLAM) [3, 5, 10]. These have been tried in several domains including reconstructing the abdominal cavity [**?**] and heart [6]. However standard SfM and SLAM assume the 3D scene is rigid, which is unrealistic during intervention. Nonrigid-SLAM extensions have been proposed, yet these require strong motion models, such as cyclic deformation [10], learned low-rank shape bases [6] or conformal surface extension [9]. Shape-from-Shading (SfS) is another passive method tried [12, 16] that exploits the relationship between geometry, pixel intensity and scene illumination. In contrast to SfM it can return dense 3D, but it currently has major limitations. These include the inability to handle surface discontinuities, and inherent unreliability due to SfS being a very weakly constrained problem.

3D Reconstruction with Distant Light Photometric Stereo. PS can be considered the generalisation of SfS to multiple light sources. In prior work, the *distant light source model* is nearly always used [2, 15, 8]. This serves as a basis for us, but is unsuitable at close-range where illumination attenuation is significant. A given point $\mathbf{q} = (u, v)$ in an image projects out into 3D according to an (unknown) depth function $z(u, v) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$. Its 3D position is given by $\mathbf{p} = \mathbf{K}^{-1}(u, v, 1)^\top z(u, v)$, where $\mathbf{K}$ denotes the camera's perspective intrinsics (which implies image distortion effects have been undone.) It is assumed that $\mathbf{p}$ is lit by $K \geq 3$ lights whose directions are given by the vectors $\mathbf{l}_k$. For an RGB camera, we have effectively three light sensors, with each channel sensitive to different parts of the light spectrum. Denote $c_i(u, v)$ to be the radiometrically corrected image intensity of the pixel for the i^{th} colour channel. In standard

distant-light PS, 3D shape at $\mathbf{p}$ is constrained by lambertian reflectance according to: $c_i(u,v) = \sum_k \mathbf{l}_k \cdot n(u,v)\,\mathbf{A}_{ik}$. Here, $n(u,v)$ denotes the surface normal. $\mathbf{A}$ is a $3 \times k$ matrix where $\mathbf{A}_{ik} \geq 0$ holds the illumination response of the i^{th} channel as a function of surface albedo and the k^{th} light's power spectrum [2]. Distant Light PS involves using these constraints to solve for $n(u,v)$. This is a small, quadratically-constrained Linear Least Squares (LLS) problem. Once estimated, dense 3D shape is recovered by integrating $n(u,v)$ in a second optimisation phase. Note however that absolute depth is not recoverable, and shape is given up to an unknown scale factor.

2 Close-Range Photometric Stereo

In this section we generalise the PS problem to handle close range light conditions. We present a new low-parameter illumination model which models very well a laparoscope's light source and give a method for quick and practical light calibration. We retain the lambertian model in this work, and handle specularities via saturation detection. This simplified model allows for tractable dense 3D reconstruction. We further advocate lambertian constraints in another important respect. By placing polarizing filters over the light and camera, specular reflection can be hugely reduced, leaving mostly only the lambertian term. Thus, with filters, the lambertian model is arguably a good one to use (for reconstruction purposes) even if the viewed surface comprises specular reflections. We start by extending the PS constraints to the following general form:

$$c_i(u,v) = l_k(p(u,v)) \cdot n(u,v)\, w_k(p(u,v))\mathbf{A}_{ik}, \quad p(u,v) = \mathbf{K}^{-1}(u,v,1)^\top z(u,v) \quad (1)$$

Here $l_k(\cdot) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is now not a constant, but a *spatially varying* light vector function. $w_k(\cdot) \in [0:1]$ is also a spatially varying function that gives the amount of light attenuation from the k^{th} light source to a point in 3D space. We say that the light model is *calibrated* if $l_k(\cdot)$ and $w_k(\cdot)$ are known. Close-range PS then involves solving the following variational least squares system:

$$\begin{aligned} \arg\min_{z(u,v)} \textstyle\int_{(u,v)\in\Omega} \sum_{k=1}^{K} \sum_{i=1}^{3} \left(l_k(p(u,v)) \cdot n(u,v) w_k(p(u,v))\mathbf{A}_{ik} - c_i(u,v)\right)^2 + \\ \lambda \textstyle\int_{(u,v)\in\Omega} \nabla z(u,v)^2 du dv \end{aligned} \quad (2)$$

Here the domain Ω denotes an image region bounded by the surface. The first line enforces the PS constraints, and the second enforces surface smoothness weighted by λ. Let us now step back and compare close-range PS to the distant-light case. Firstly (2) cannot be broken down into two convex problems (normal estimation, followed by depth estimation). This is because the PS constraints depend on both depths and normals. As such, it is a harder optimisation problem. However, it is the fact that (2) depends on depths that allows us to recover *absolute* distances to the camera (in mm), unlike distant-light PS.

2.1 Illumination Modelling

We now turn to modelling and calibrating the light source functions $l_k(\cdot)$ and $w_k(\cdot)$. Our goal is to have accurate models, yet which can be calibrated easily and

optimally. We propose using an attenuating point light source, with a bivariate polynomial which can model light fall-off caused by both distance and angular attenuation. This is a flexible model and a generalisation of the inverse-squared fall-off model [16], which we have found to be rather poor. The model's parameters comprise firstly a light source centre: $\mathbf{u}_k \in \mathbb{R}^3$. The illumination vector at any 3D point $\mathbf{p}$ is given by the unit direction $l_k(\mathbf{p}) = (\mathbf{p} - \mathbf{u}_k)/||\mathbf{p} - \mathbf{u}_k||_2$. The attenuation function is a joint function of the distance from $\mathbf{p}$ to $\mathbf{u}_k$: $d(\mathbf{p}, \mathbf{u}_k) = \|\mathbf{p} - \mathbf{u}_k\|_2$, and the angular attenuation w.r.t. the light's *principal direction* $\mathbf{v}_k$: $\psi(\mathbf{p}) = \angle(l_k(\mathbf{p}), \mathbf{v}_k)$. This angular attenuation is important to model the spotlight characteristics of the light source. Here $\angle(\cdot\cdot)$ denotes taking the angle between two vectors. The attenuation function then writes as:

$$w_k^{-1}(\mathbf{p}) = \sum_{s=1}^{S} \sum_{t=1}^{T} \mathbf{W}_{st}^k d(\mathbf{p}, \mathbf{u}_k)^s \psi(\mathbf{p}, \mathbf{v}_k)^t \tag{3}$$

Here $\mathbf{W}^k$ holds the k^{th} light's polynomial coefficients up to order (S, T).

2.2 Light Calibration

Light calibration involves finding, for each coloured light source, the values $\{\mathbf{u}_k, \mathbf{v}_k, \mathbf{W}_k\}$. Typically for endoscopes the light sources remain rigidly fixed in the camera's coordinate frame, which means that calibration can be done in a one-time offline process. We divide the calibration problem into first determining the light centres $\{\mathbf{u}_k\}$, and then using these to determine the attenuation terms $\{\mathbf{v}_k, \mathbf{W}_k\}$. This 2-stage approach gives a convex solution to light calibration, and so global optimality is guaranteed. The light centres $\mathbf{u}_k$ can be found easily by detecting and triangulating their positions on a reflective calibration target [14] (Fig. 2(a)). To calibrate $\{\mathbf{v}_k, \mathbf{W}_k\}$ we use Eq. (1) to optimise these terms using ground-truth training samples. The data is acquired using images of a diffuse planar calibration target with a checker pattern printed on one side (Fig. 2(b)). The pattern gives us the plane's pose in each image. For each colour filter, we gather a large set of training samples $\{(c_r, c_g, c_b, \mathbf{n}, \mathbf{p})\}$. Now, for a given value of $\mathbf{v}_k$, $\mathbf{W}_k$ can be optimised via LLS. We thus calibrate by densely sampling $\mathbf{v}_k$ over the unit sphere, solving for $\mathbf{W}_k$, and retaining the solution with minimal least-squares error w.r.t. Eq. (1). We can select the best (S, T) by minimising the fitting error on a separate validation set. In practice it is usually unnecessary to go beyond 4^{th} order.

2.3 Reflectance Model Learning

For any tissue we wish to recover we also need an estimate of $\mathbf{A}$. For $K = 3$ lights this is only a small 3×3 matrix and can be determined with training data. There are two main approaches one could take for this. The first is to learn $\mathbf{A}$ prior to intervention for a range of tissue classes. The second is to assume the training data can be acquired during intervention by some other means. Currently we adopt this second approach, and place a small marker on the tissue (e.g. Fig. 4 (a-e)). By computing the marker's pose, and sampling tissue intenities close to the marker, we have the necessary data to compute $\mathbf{A}$ from Eq. (1).

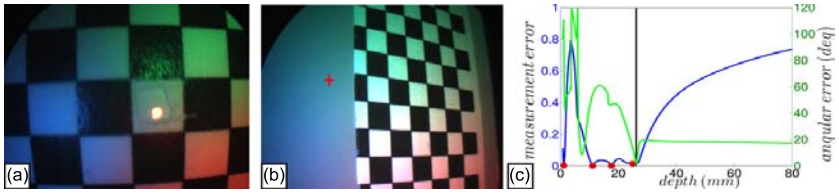

Fig. 2: Calibration of light source centres using specular reflection (a) and light source attenuation using views of a lambertian planar target (b). For the pixel marked at the cross, there may exist multiple depths which locally satisfy the close-range PS constraints.

3 3D Reconstruction

The 3D reconstruction problem involves solving (2) which is a challenging non-convex problem. Here we present an effective 2-stage approach to find a good minimum. In stage 1 we first solve for depth at each pixel *locally* using only that pixel's colour information. When computed in isolation from other pixels these solutions are however usually non-unique. In stage 2 these local solutions are then resolved by solving depth *globally* across the image. At any pixel (u, v), the intensities $c_r(u, v), c_g(u, v), c_b(u, v)$ provide us with 3 constraints on shape according to Eq. (1). Locally, we have three unknowns, one for $z(u, v)$ and two for $n(u, v)$. This is a polynomial optimisation problem whose number of solutions depends foremost on the order of $w_k(\cdot)$. We propose a fast method to find these solutions as follows. We regularly sample depth in the range $\tilde{z}(u, v) \in [0 : z_{max}]$, where z_{max} denotes the maximum working distance of the laparoscope (typically 150mm). Using Eq. (1), each sample is used to solve for a putative surface normal $\tilde{n}(u, v)$, which is a small convex sub-problem. We can then evaluate the solution pair $(\tilde{n}(u, v), \tilde{z}(u, v))$ against the measured intensities predicted by Eq. (1), and retain the solutions which are optimal. We illustrate this approach in Fig. 2(b-c). In Fig. 2(b) a planar surface is 34.5mm from the camera's optical centre. For the pixel marked at the cross, we show in 2(c) on the x-axis the depth along the pixel's viewing ray. In green we show the angular error of the surface normal estimated at that depth using Eq. (1). In blue we plot the prediction error of the pixel's intensity. Clearly, there is a 4-fold solution ambiguity, marked by red dots with zero prediction error. The rightmost solution is closest to the true depth, marked by a black line.

The sets of solutions computed at each pixel can be resolved in a second process by enforcing surface continuity between pixels. This can be achieved by constructing a Markov Random Field (MRF), whose nodes correspond to pixels and edges connect neighbouring pixels. These edges enforce consistency between pixels' normal and depth estimates, and form a graph with sub-modular pairwise interaction terms. We have found the MRF's energy can be minimised well with belief propagation and the solution gives us a reasonable initial solution to the depth map. This is then refined by optimising the original problem (2). In practice we discretised Ω using the pixel grid and iteratively refine $z(u, v)$ with nonlinear Gauss-Newton iterations.

4 Experimental Results

Ex-Vivo Experimentation. We have tested our approach ex-vivo using two organs; a section of pig liver and a pig kidney. We have performed ground truth evaluation by first scanning these surfaces with a Structured Light Scanner (SLS). In Fig. 3(a,g) we show the kidney and liver ground truth surfaces. To learn the organs' reflectance models, we attached a small planar checker marker to the organ to give us depth and normal information (Fig. 3(b,h)), and used the non-specular tissue colour around the marker as the intensity training data. We handled external laparoscope tracking using a mounted calibration target, giving us the coordinate transform from the laparoscope's view to the 3D SLS surface. We then imaged the organs with the laparoscope at varying positions (Fig. 3(c,d,i,j)). For each image, we manually segmented the organ from the background to obtain Ω. In these experiments we did not use polarizing filters and specularities were handled with simple methods by detection based on colour saturation. For any specular region, its pixel data does not contribute to the first term in Eq. (2). 3D reconstruction was then performed. With our current Matlab implementation this takes approximately 30s to process an image. However, much can be parallelised so a GPU implementation would be significantly faster. We used the same value of λ for all images (which was set by hand) and measured the absolute error in depth against ground truth. We show the results for the four images below in Fig. 3(e,f,k,l). In general the surfaces are reconstructed quite faithfully. Greater errors occur towards some boundaries of the surface, which is due to surface inter-reflection from the background and slight mis-alignments of the ground truth scan. Note in Fig. 3(l) the larger error occurs at a region where the red channel becomes saturated, which corresponds to losing a PS constraint.

In-Vivo Experimentation. We have also obtained some preliminary in-vivo experimental results by testing reconstruction on the liver of a live pig under anesthetic. To acquire ground truth data our surgeon placed 5 4.5mm wide checker-markers on the liver using surgical graspers[1]. This gives a sparse set of ground truth depths and surface normals. Fig. 4(a-e) show a selection of images of the markers taken by the laparoscope. We used one of these markers to learn in-vivo the liver's surface reflectance model by sampling pixel intensities at tissue locally surrounding the marker. To perform reconstruction, we took the image domains Ω to be the entire image, but excluding the marker locations. The 3D reconstructions are shown in Fig. 4(f-j), each corresponding to the input image shown above it. We also render the laparoscope's tip, indicating its absolute distance to the surface. In green we mark the predicted surface normal at each marker, computed from the gradient of the reconstructed surface. In total, we have performed reconstruction for 30 images. Quantitative performance has been studied by measuring the error in the predicted depths and normals of the markers. In Fig. 4(k-l) we show the error distribution in both depth (in mm) and surface orientation (in degrees).

[1] We thank Dr. Revaz Botchorishvili for his kind help in acquiring the in-vivo data.

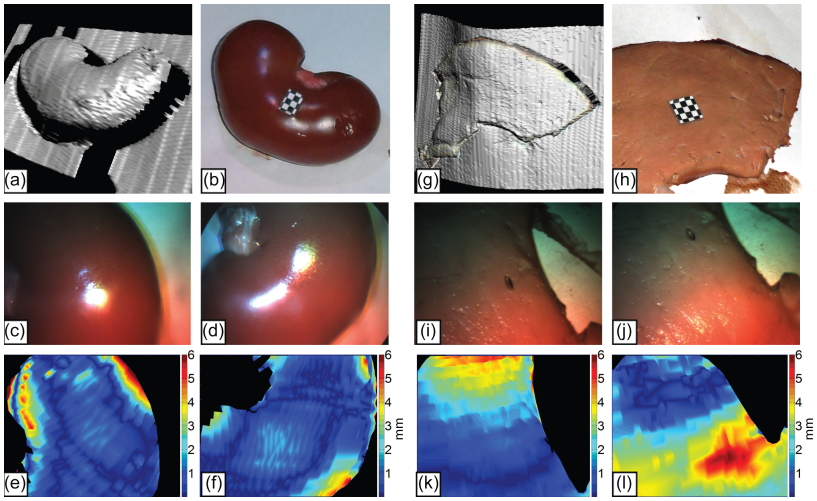

Fig. 3: Ex-vivo experimental validation of close-range PS

5 Conclusion and Future Work

In this paper we have aimed to answer an important open question: can PS be used to successfully reconstruct surfaces in the close-range conditions of MIS? Our preliminary results suggest yes. Focusing on laparoscopy we have extended distant light PS to handle short-range lights, developed methods for calibration, and an optimisation framework to achieve good solutions to depth. In contrast to other active methods tried in MIS, our approach can be used with an existing laparoscope with only minor modification. Unlike SfM, the approach handles textureless surfaces and does not require motion constraints. Unlike SfS, the method is stable and we can compute absolute depth. There is still further research to be done before it can handle unconstrained clinical conditions. Open challenges include handling spatially varying reflectance, handling time-varying illumination caused by changes in brightness or exposure, to learn different reflectance classes *a priori*, to handle tool occlusions, and to handle depth discontinuities with robust smoothing priors. We will also investigate how the recovered 3D can help solve the challenging problem of pre-operative/intra-operative registration.

References

1. J. D. Ackerman, K. Keller, and H. Fuchs. Surface reconstruction of abdominal organs using laparoscopic structured light for augmented reality. *3DICA*, 2002.
2. G. Brostow, C. Hernandez, G. Vogiatzis, and R. Stenger, B. andCipolla. Video normals from colored lights. *PAMI*, 2012.
3. D. Burschka, M. Li, M. Ishii, R. H. Taylor, and G. D. Hager. Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery. In *MICCAI*, 2004.

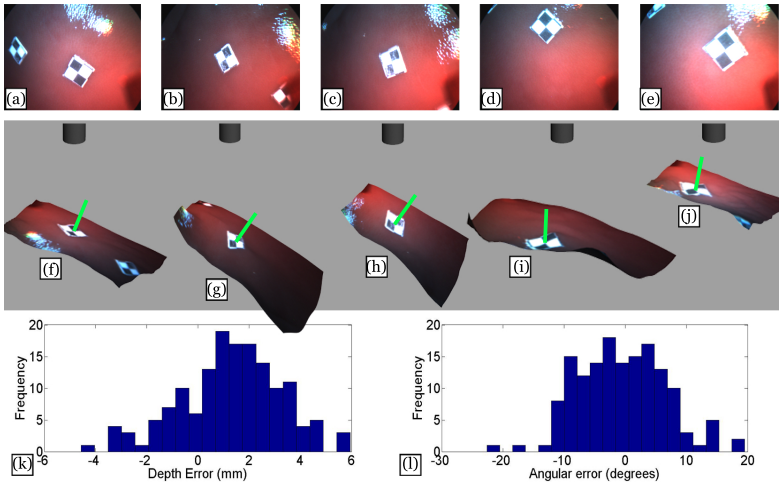

Fig. 4: In-vivo experimental validation of close-range PS

4. F. Devernay, F. Mourgues, and È. Coste-Manière. Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery. In *MIAR Workshop*, 2001.
5. M. Hu, G. Penney, P. Edwards, M. Figl, and D. Hawkes. 3D reconstruction of internal organ surfaces for minimal invasive surgery. In *MICCAI*, 2007.
6. M. Hu, G. Penney, D. Rueckert, P. Edwards, F. Bello, R. Casula, and D. Figl M.and Hawkes. Non-rigid reconstruction of the beating heart surface for minimally invasive cardiac surgery. In *MICCAI*, 2009.
7. W. W. Lau, N. A. Ramey, J. J. Corso, N. V. Thakor, and G. D. Hager. Stereo-based endoscopic tracking of cardiac surface deformation. In *MICCAI*, 2004.
8. J. Lim, J. Ho, M.H Yang, and D. Kriegman. Passive photometric stereo from motion. *ICCV*, 2005.
9. A. Malti, A. Bartoli, and T. Collins. Template-based conformal shape-from-motion from registered laparoscopic images. In *MIUA*, 2011.
10. P. Mountney and G. Z. Yang. Motion compensated SLAM for image guided surgery. In *MICCAI*, 2010.
11. J. Penne, K. Höller, M. Stürmer, T. Schrauder, A. Schneider, R. Engelbrecht, H. Feußner, B. Schmauss, and J. Hornegger. Time-of-flight 3D endoscopy. In *MICCAI*, 2009.
12. C. H. Quartucci Forster and C. L. Tozzi. Towards 3d reconstruction of endoscope images using shape from shading. *GPI*, 2000.
13. D. Stoyanov, A. Darzi, and G.Z. Yang. Dense 3D depth recovery for soft tissue deformation during robotic assisted laparoscopic surgery. In *MICCAI*, 2004.
14. D. Stoyanov, D. Elson, and G.Z. Yang. Illumination position estimation for 3D soft-tissue reconstruction in robotic MIS. In *IROS*, 2009.
15. R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 1980.
16. C. Wu, S.G. Narasimhan, and B. Jaramaz. A multi-image shape-from-shading framework for near-lighting perspective endoscopes. *IJCV*, 2009.