

Dense Surface Reconstruction With Shadows in MIS

Bingxiong Lin, *Student Member, IEEE*, Yu Sun, *Senior Member, IEEE*, and Xiaoning Qian, *Member, IEEE*

Abstract—3D reconstruction of internal organ surfaces provides useful information for better control and guidance of the operations of surgical tools for minimally invasive surgery (MIS). The current reconstruction techniques using stereo cameras are still challenging due to the difficulties in correspondence matching in MIS, since there is very limited texture but significant specular reflection on organ surfaces. This paper proposes a new approach to overcome the problem by introducing weakly structured light—actively casting surgical tool shadows on organ surfaces. The contribution of this paper is two-fold: first, we propose a robust approach to extract shadow edges from a sequence of shadowed images; second, we develop a novel field surface interpolation (FSI) approach to obtain an accurate and dense disparity map. Our approach does not rely on texture information and is able to reconstruct accurate 3D information by exploiting shadows from surgical tools. One advantage is that the point correspondences are directly calculated and no explicit stereo matching is required, which ensures the efficiency of the method. Another advantage is the minimum hardware requirement because only stereo cameras and a separated single-point light source are required. We evaluated the proposed approach using both phantom models and *ex vivo* images. Based on the experimental results, we achieved the precision of the recovered 3D surfaces within 0.7mm for phantom models and 1.2mm for *ex vivo* images. The comparison of disparity maps indicates that with the addition of shadows, the proposed method significantly outperforms the state-of-the-art stereo algorithms for MIS.

Index Terms—Dense surface reconstruction, low texture, minimally invasive surgery, stereo reconstruction, weakly structured light

I. INTRODUCTION

DUE to the benefits of minimized trauma, shorter hospitalizations, and lower infection risk, minimally invasive surgery (MIS) has been considered an alternative to open-cavity surgery. However, there are still many challenges in the current MIS systems, such as narrow field of view and incapability of capturing and displaying depth. The real depth information of the actual internal organs and the surgical scene is valuable to surgeons and can potentially be used with

many cutting-edge computer aided interventions such as 3D surgery visualization and planning. Other benefits of the intra-operative depth information is discussed in [1], [2]. Recently, there has been an active research interest in how to recover 3D surfaces of low-texture organs. In [3], constraint-based factorization method of structure from motion was used to reconstruct the 3D structure from endoscopic video. A couple of methods with a focus on cardiac surgery have been proposed to estimate the depth information by tracking feature points on a surface [4]–[6]. However, specific geometric models are usually assumed in order to use the sparse feature points to estimate the dense surface. For example, B-Splines were used to model the surface in [4] and Thin-Plate Splines (TPS) were assumed in [5], [6].

On the other hand, stereo reconstruction was considered one of the most practical ways to recover depth for MIS, since no extra sensors were required [1]. As stereo imaging systems for MIS become more popular, there is more research on how to implement stereo reconstruction. Stereo reconstruction has been a classic method to recover depth information and is able to provide dense reconstruction when a robust correspondence matching is performed on a scene with enough distinguishable texture. Reviews of traditional stereo reconstruction can be found in [7], [8]. However, the surfaces of most organs inside the abdomen do not have rich distinguishable texture, and the wet, shiny, and curved surface creates broad specular reflection, both of which make stereo reconstruction very difficult. To solve these problems, Stoyanov et al. [1] proposed to start with a sparse set of feature point correspondences and propagate disparity information using information from nearby regions. Later, Stoyanov’s method was further developed and combined with sparse simultaneous localization and mapping (SLAM) to create a dense tissue model in [2]. The method continues to update new images with the existing model and is able to dynamically expand the field of view of a laparoscope. However, Stoyanov’s method has difficulty in low texture areas, because propagation becomes difficult when there is not enough texture available.

To recover a surface with low texture, some researchers aimed to actively project patterns on the tissue surfaces. Wu et al. developed imaging systems which projected grid patterns [9] or laser strips [10] to reconstruct the 3D structure of cervix and assist the diagnosis of cervical cancer. Fuchs’s team [11] designed and implemented a miniature projector that projected structured stripe patterns on the abdominal organs. However, the stripe patterns are distractive, and the projector requires special insertion ports, which would increase surgery difficulty and time. Instead of relying on extravagant stripe patterns, we

Manuscript received September 19, 2012; revised January 02, 2013; revised March 22, 2013. This material is based upon work supported by the National Science Foundation under Grant No. 1035594.

Bingxiong Lin is with the Department of Computer Science and Engineering, University of South Florida, Tampa, FL., 33613 USA (e-mail: bingxiong@mail.usf.edu)

Yu Sun is with the Department of Computer Science and Engineering, University of South Florida, Tampa, FL., 33613 USA (e-mail: yusun@cse.usf.edu)

Xiaoning Qian is with the Department of Computer Science and Engineering, University of South Florida, Tampa, FL., 33613 USA (e-mail: xqian@cse.usf.edu)

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

observed that during the surgical process, shadows generated by a surgical tool could provide a weak but structured pattern, which gives a cue to generate “correspondences.”

It is worth noting that, not surprisingly, both in MIS and computer visualization, researchers have noticed that shadows can significantly improve depth perception [12]–[14]. The study of how to generate optimum shadows in terms of contrast and location of shadow-casting illumination by using a second endoscope was introduced in [12]. A secondary light source was also used in [13] to carefully cast an “invisible shadow,” which was digitally detected and enhanced to provide a depth cue. It should be noted that in order for the cameras to capture the shadows cast by surgical tools, the cameras and light source should be separated, which is also adopted in this paper.

To the best of our knowledge, the first work to use actively-cast shadows to recover low texture surface was introduced in [15] with a method called “weakly structured light” [15]. However, that method required a calibrated light source and placed two perpendicular planes in the scene. These two requirements are difficult to be satisfied, as the space is very limited in an MIS environment. In this paper, we remove these two requirements by using stereo cameras and a separated light source. Our method first extracts the shadow borders and interpolates them with epipolar lines to generate disparity maps. Other than being able to achieve dense and accurate reconstruction results, this approach does not require stereo matching, which is much more computing-intensive than shadow extraction in the proposed method. Therefore, we expect that this method could be much more efficient than the traditional stereo-matching-based approaches with an optimized implementation. Another advantage of our method is that only stereo cameras and a separated light source are required, since surgical tools are part of a standard MIS setup and surgeons wave surgical tools in front of organs already. It should be noted that our method only recovers a relatively small area of tissue surface at one time due to the narrow field-of-view in MIS as noted in [1]. To overcome this limitation, as proposed in [2], camera localization using SLAM technique can be integrated to combine small tissue surface patches recovered at different time and obtain a larger recovered 3D tissue surface. The SLAM technique is beyond the scope of this paper and will not be discussed here. We have evaluated the proposed approach on different phantoms and *ex vivo* organs and report the accuracies of reconstructed surfaces in comparison with state-of-the-art algorithms.

II. METHODS

A. System Overview

We propose to use weakly structured light to recover the dense 3D surfaces of internal organs with stereo cameras. Our method does not require a projector or laser stripe. Instead, similar to [15], we actively cast shadows on the object as a cue to establish semi-dense stereo correspondences. There are four major steps involved: shadow curve extraction, intersection of curves and epipolar lines, field surface interpolation (FSI), and 3D reconstruction.

First, a series of images containing shadows is obtained. The shadow boundaries are extracted and used as shadow

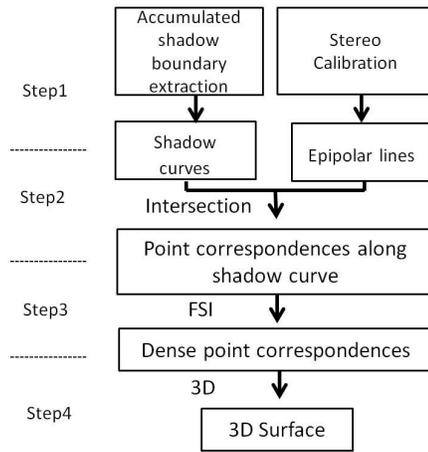


Fig. 1. Outline of the proposed method with four major steps.

curve correspondences between the two corresponding images. Then, epipolar lines are calculated and used to intersect with the shadow curves to efficiently generate precise point correspondences along the curve pair from two images. The accuracy of point correspondences is further improved to sub-pixel accuracy by proper interpolation. Finally, we develop a novel FSI approach to estimate the points that are between two shadow curves by exploiting both the spatial and stereo calibration information to generate dense correspondences between two images, which are used to recover the organ surfaces. The overall scheme of our approach is illustrated in Fig. 1.

B. Shadow Curves Extraction

Since the accuracy of shadow extraction directly affects the accuracy of the recovered surface, it is important to extract the shadow borders in both images as precisely as possible. Our shadow extraction method is based on two assumptions. Firstly, the scene is stationary during the shadow casting process. The static scene is also required in [15], which processes shadows on the temporal domain. Secondly, in order for the surfaces to clearly display the shadow boundaries, we assume the surfaces are locally smooth. It should be noted that this is a relatively weak assumption and most tissue surfaces are locally smooth. In fact, locally smooth surface is also necessary for structured light based 3D reconstruction methods to project clear patterns. Besides these assumptions, it is worth noting that our method can not extract shadow boundaries from self-shadowed areas, because the intensity changes are very small in those areas. This is an inherent limitation of methods using shadow for 3D reconstruction, such as [15].

Even though the temporal shadow edge has been used to estimate the shadow time for each pixel and has been shown to be very accurate [15], some of its limitations prevent it from being used as it is. For example, that method has difficulty in the self-shadow area. Also, it assumes that the shadow moves forward only, specifically from left to right. This is an unreasonable requirement, because human hands may, at times, be shaky, which makes the shadow move back and forth

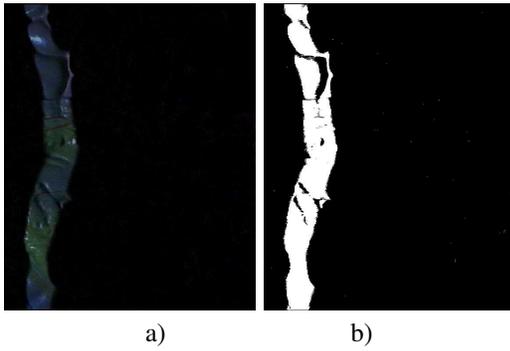


Fig. 2. a) Example of a difference image. b) Shadow mask image with shadow area shown as white. The image is taken on an intestine phantom, whose surface is uneven and its anatomical structure is better displayed in Fig. 5 and 10b.

and causes the algorithm to become unstable. Our method is designed to overcome these problems.

In [16], Agrawal introduced a way to detect depth edges and shadow edges with multi-flash light sources. It has been shown that the method is effective for handling self-shadows. In this paper, a sequence of images with a moving shadow rather than a fixed shadow is used. Similar to [16], a shadow-free image is generated by taking the maximum of intensity value at every pixel from the sequence of images, which is called the reference image I_{ref} . A difference image is defined as:

$$I_{diff} = I_{ref} - I, \quad (1)$$

where I_{diff} reflects the intensity changes of the pixels with and without shadows, which is exactly the main property of the shadow area. One example of a difference image is shown in Fig. 2(a). Based on the difference image, adaptive thresholds are set for different rows to discriminate shadow areas from one another. Similar to [15], we calculate maximum and minimum intensity for pixels along each row in each image, and the mean value is used as the threshold for each row. The shadow mask is defined in the following equation:

$$I_{mask} = I_{diff} > threshold. \quad (2)$$

The above method naturally marks the shadow area white and the other area black, as shown in Fig. 2(b), as the intensity change of the shadowed area is much larger than the other areas. Due to the existence of noise in the camera sensors and changes of reflectance, there could be isolated sparse white dots distributed within the dark region, which can be easily filtered out with a median filter.

In practice, only one shadow scan is enough for 3D reconstruction and the shadow is scanned along one direction. We observe that during the shadow scan, the shadowed area increases gradually and stably on the locally smooth surfaces. Therefore, we propose to accumulate the shadow area and extract the rightmost border as the shadow curve. The binary accumulated shadow image is initialized as a black image. Its formal definition is given iteratively as in the following equation, where the operation is pixel-wise.

$$I_{acc} = Max(I_{acc}, I_{mask}). \quad (3)$$

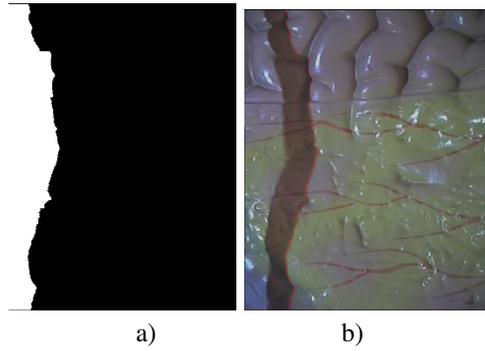


Fig. 3. a) One example of an accumulated shadow mask image; b) Image with detected shadow border overlaid. Resolution of the raw image is 640*480. The camera-to-target distance is from 11cm to 15cm. More details are available in “Experiments and Results” section.

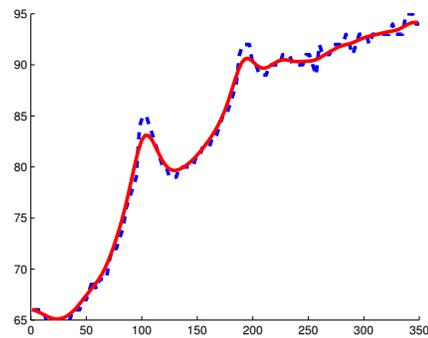


Fig. 4. Shadow curves before and after LWR. Blue dotted line represents the original shadow curve. Red solid line represents the curve after LWR.

It is worth noting that only the newly-generated shadow areas are processed and the backward shadows will be ignored. This makes shadow extraction more robust and solves the potential shaky hand problem. One example of an accumulated shadow mask image is shown in Fig. 3.

Intuitively, the shadow curve is defined along the vertical direction as the rightmost border of the accumulated region. For each row, the rightmost column of shadow is recorded. Due to the camera speed, resolution and the tool motion, the shadow boundaries in the image might be blurry. For the blurry shadow boundary, the shadow curve is not unique and depends on the threshold value. The shadow curve after thresholding is typically zigzagging due to the discretization nature of image. In addition, the curve is highly sensitive to the image noise, which makes the shadow curve in the left image do not correspond to the curve in the right. Because the surface is assumed to be locally smooth, the shadow curve is expected to be locally smooth. We apply LWR to smooth the zigzagging curve locally, which makes the left and right curves more consistent and robust towards the image noise. Since the curve might contain multiple segments, the locality is extended to 2D image space so that each segment can be smoothed separately. After LWR, the coordinates of curve pixels reach sub-pixel accuracy. A shadow border before and after LWR is shown in Fig. 4.

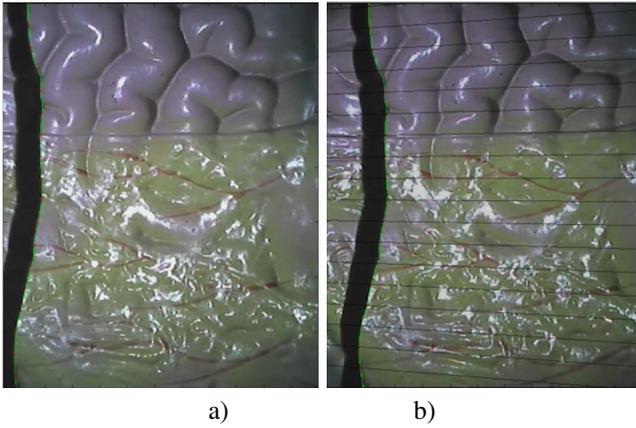


Fig. 5. Illustration of intersection of shadow curve and epipolar lines. a) Left image. b) Intersection of shadow curve and epipolar lines on right image. The shadow is casted vertically on the image so that its intersection with the horizontal epipolar lines is unique.

C. Shadow Curves Intersection

After the shadow curves are obtained, for points along a curve in one image, we find their corresponding points in the other image by using the intersection between the shadow curves and epipolar lines. When the images are not rectified, the epipolar lines can be calculated using fundamental matrix from the calibration results. When the images are rectified, the epipolar lines are just the image rows. For each point in the left image, different from the traditional stereo matching method that searches along the epipolar line in the right image, our approach directly calculates the intersections between the shadow curves and epipolar lines, as proposed in [17]. Since the corresponding point in the right image should lie on both the shadow curve and the epipolar line, their intersection point is exactly the corresponding point. This is illustrated in Fig. 5. To simplify the problem, we arrange the two cameras perpendicular to the tool so that the epipolar lines are perpendicular to most of the casted shadows and there will be a unique one intersection. However, there are extreme and rare cases, in which there might be more than one intersection from zigzags by discretization. For those cases, we use the order of the intersections on the epipolar line to define the matching. On the other hand, the shadow curves in the self-shadow areas are marked as invalid and there will be no intersections, which is the reason why no corresponding points can be found in the self-shadow areas. For computation efficiency, even considering the overhead of the shadow boundary extraction, direct calculation of the intersection point should be much more efficient than stereo matching, which simply requires an extra 1D search for each pixel pair.

D. Field Surface Interpolation

The 3D coordinates of the pixels on the shadow curves can be directly calculated by the traditional triangulation method [18] or from the disparity values. Those shadow curves divide the image into small regions. For pixels inside of those regions, their 3D coordinates can be interpolated by nearby pixels whose depths have been calculated. The interpolation method used here should exploit two constraints: spatial constraint

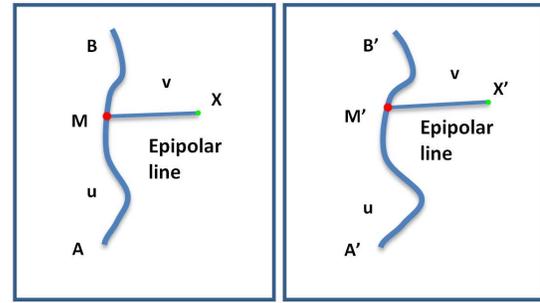


Fig. 6. Mapping defined by a pair of curves and epipolar lines.

and stereo constraint. The spatial constraint is based on the fact that the pixel is between two shadow curves. The stereo constraint comes from the stereo calibration. Bouget [15] proposed to estimate the shadow time for each pixel, which could not take into consideration of stereo information. In the surface reconstruction community, an interpolation in 3D space is always used, such as Delaunay triangulation, does not consider the stereo information. Here, we propose a novel FSI method, which incorporates both spatial information and stereo calibration information.

First, consider only a single pair of curves on two images. It is known that a pair of lines, one from each image, can define a mapping between the coordinates of the two images [19]. The difference of a curve and a line here is that each point on a line has the same normal, while different points on a curve may have different normal directions that might intersect with one another, which makes the mapping between two images not bijective. To avoid the intersection of normals, epipolar lines provide a natural alternative, which are guaranteed to have no intersection. Specifically, each point on the curve is attached to a direction that is along the corresponding epipolar line. Now, each pixel on the curve has two coordinates: one is along the epipolar line and the other is along the curve itself. The mapping is illustrated in Fig. 6.

Curve AB in the first image corresponds to curve $A'B'$ in the second image. For each point X in the first image, its epipolar line intersects with curve AB at M . The corresponding epipolar line in the second image intersects with curve $A'B'$ at M' . For X , its coordinate along MX is defined as v , which is calculated as follows:

$$v = \frac{|MX|}{|\widehat{AB}|}, \quad (4)$$

where $|\widehat{AB}|$ represents the arc length of curve AB . In the second image, the same v is used as the coordinate along $M'X'$ to find X' . That is,

$$|M'X'| = v * |\widehat{A'B'}|. \quad (5)$$

After the above steps, for each point in the first image, a unique point in the second image is found. Also, each point in the second image corresponds to a unique one in the first image. This gives a bijective mapping.

The mapping defined above is only for the special case with one pair of curves. In practice, a large number of pairs of shadow curves are available. For general case, those curves

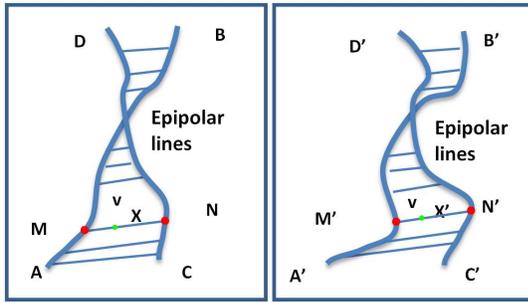


Fig. 7. Mapping defined by two pairs of curves and epipolar lines. The above figure shows the case with intersecting curves.

might intersect with each other and divide the image into small regions. Instead of taking a global mapping, a local mapping is defined for each region. Even though a region might be surrounded by multiple curves, for simplicity only the mapping for the region surrounded by two curves is explained, as illustrated in Fig. 7. For the region surrounded by curve AB and curve CD in the first image, each point X lies on one epipolar line that intersects with curve AB at M , curve CD at N . The coordinate of X along segment MN is defined as:

$$v = |MX|/|MN|. \quad (6)$$

As in the single pair case, the corresponding point X' is defined as the point on $M'N'$, which has coordinate v :

$$|M'X'| = v * |M'N'|. \quad (7)$$

It is worth noting that the mapping for all pixels we defined here is consistent with the mapping for pixels on curve boundaries.

E. 3D Reconstruction

The derived mapping gives dense correspondences between the two images. The proposed method establishes dense correspondences and depends only on the information of the shadow curves and the epipolar lines. This means that no texture on the object surface is used. The 3D reconstruction can be performed with or without image rectification. Most stereo reconstruction methods perform rectification before stereo matching, which simplifies the 2D correspondence matching into a 1D search task. After rectification, the 3D reconstruction is equivalent as building the disparity map and disparity values have been chosen as the standard for the comparison of different stereo matching algorithms [7]. Following the same framework for the comparison purpose, we also perform rectification and build disparity map. For accuracy and efficiency, we adopt the rectification from [20]. It is worth noting that our method is not limited to rectified images.

III. EXPERIMENTS AND RESULTS

To take advantage of our approach, it is necessary to have stereo cameras and a separated light source. In a regular MIS, a stereoscope can be used along with a light source through a separate port for the generation of natural shadows, as in [12],

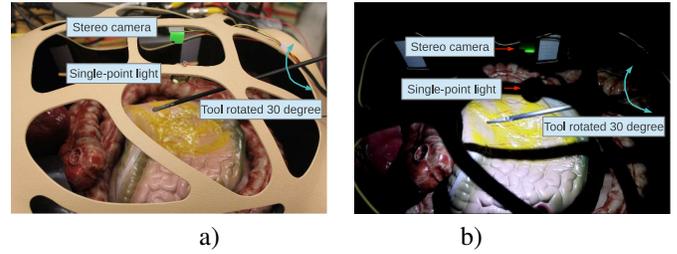


Fig. 8. a) Experiment setup with a stereo camera, a single-point light source and a surgical tool. b) Illustration of how the shadow is casted by waving the tool in front of the light.

[21]. It is also possible to use a new shadow telescope [21] with light delivered through a separate illumination cannula. This approach naturally fits with our novel wireless camera setup, as introduced in [22].

The experiment setup for this paper is illustrated in Fig. 8, which contains a rigid shell with an insufflated abdomen (Chamberlain Group, MA, USA). The cameras we used are micro wireless CCTV cameras (10mm diameter), with 640×480 resolution and 30 fps speed. The cameras were synchronized by a SENSORAY frame grabber. The light source was built from a Cree XLamp XM-L LED with a footprint of $5mm \times 5mm$. This single LED can deliver up to 1000 lumens. The abdomen has size of about $39cm \times 34cm \times 21cm$ (length \times width \times height). The camera-to-target distance ranges from 11cm to 15cm. Each camera's field of view covers area of size about $10cm \times 9cm$ and their overlap field of view has size of about $7cm \times 9cm$. The distance between the stereo camera and the single-point light source is about 6cm. The surgical tool has diameter of 5mm and length of 34cm. To cast the shadow, the surgical tool is inserted in the abdomen and horizontally rotated in front of the light. The perpendicular distance of the tool to the light source is about 7cm and 8cm to the cameras. During the shadow casting process, the distance of the tip of the tool to the object is within 4cm-7cm. With only about 30 degrees of surgical-tool waving, the casted shadow is able to cover both cameras' fields of view. Since the waving movement is small, the motion can be achieved in most abdominal MIS surgeries. The video of the shadow casting process and the videos captured by stereo cameras are all available online (<http://rpal.cse.usf.edu/project1/index.html>).

To better illustrate the setup we used, a diagram is presented in Fig. 9. As shown in the figure, a stereo rig and a single-point light source were both mounted using needles [22], [23] on the abdominal wall. However, our setup is flexible, especially the placements of the light and cameras. This flexibility allows the light and cameras to be mounted at different positions for different surgeries.

A. Phantom and ex vivo Images

To validate the proposed method, we used the above setup to capture images and tested the algorithm on four phantoms with different types of material: a flat textured paper, an intestine, a lung, and a heart. The flat textured paper was placed on a flat board. The intestine and lung are plastic and the heart is made of silicon. To be clear, the heart phantom was used

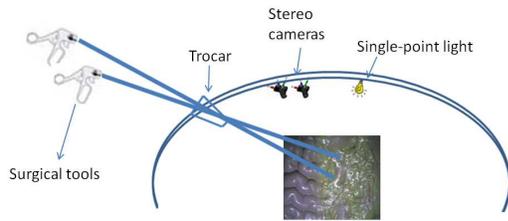


Fig. 9. Illustration of the experiment setup.

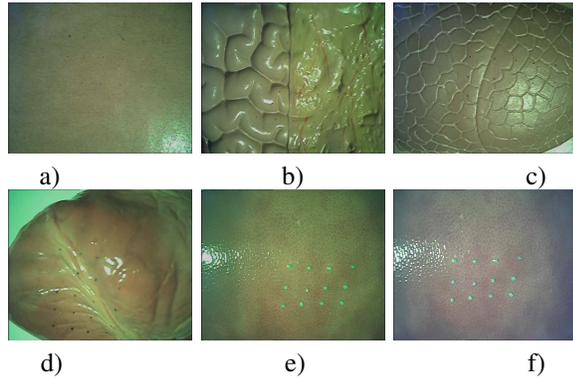


Fig. 10. a) Flat plane. b) Plastic intestine model. c) Plastic lung model. d) Silicon heart model. e) Left image of porcine liver. f) Right image of porcine liver.

only as an example for its life-like surface. We do not claim that our current approach can be used in cardiac surgery since the real heart has fast and complex motion [6]. Examples of the original images are shown in Fig. 10. It can be seen from the images that they all have the specular reflection problem. Because specular reflection is perspective-dependent, the specular reflection areas of the two cameras are different, which means correspondences based on the specular reflection texture will not be correct. Meanwhile, the texture on the images tends to be uniform and not distinctive enough, which makes it difficult to establish correspondences.

To show the performance of our method on *ex vivo* images, we tested the algorithm on images taken from a porcine liver. Because the porcine liver was wet, specular reflection and inter-reflection became more severe and caused a larger error in shadow extraction. The numerical results of both the phantoms and *ex vivo* images are available in Section IIID.

B. Disparity Maps

To illustrate the benefits of using shadow information for 3D reconstruction for MIS, we have compared our approach with traditional stereo algorithms, in which stereo cameras were calibrated and images are rectified [24]. The rectification we adopted in this paper is from [20] due to its accuracy and efficiency. After rectification, the focal length of the two cameras was 798.40 and the baseline was 11.16mm. The valid disparity range for our setup is [60 130], which is used in stereo matching algorithms as a priori. The proposed method is compared with three popular stereo matching algorithms. The first one is considered to be the state-of-the-art stereo matching algorithm applied in MIS [1], which is referred as

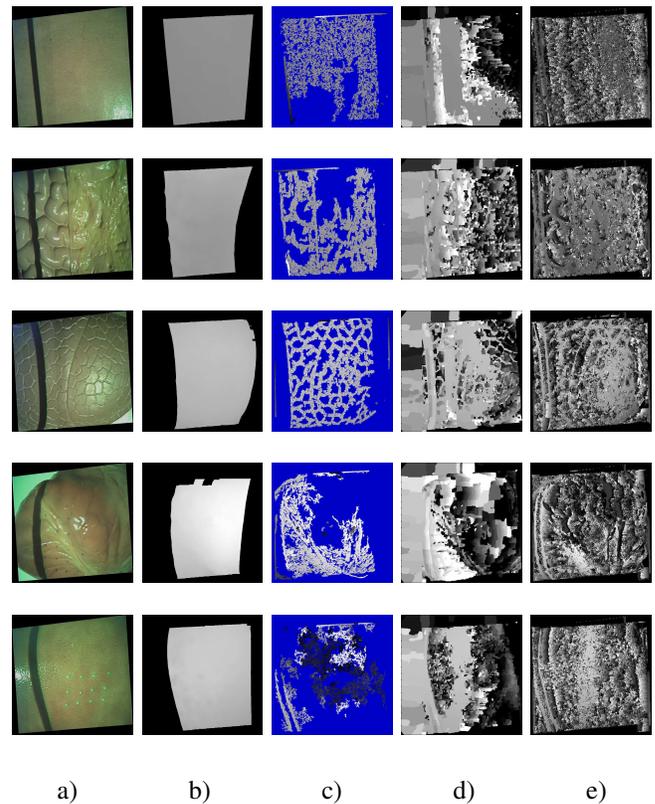


Fig. 11. a) Rectified left image with shadow and detected border. b) Disparity maps derived by our proposed method. c) Disparity maps by SP [1]. d) Disparity maps by BP [25]. e) Disparity maps by RT [26].

seed propagation (SP). The second one is referred as believe propagation (BP) [25]. Following the notation in [1], the last one is abbreviated as RT [26].

The disparity maps obtained by different algorithms are shown in Fig. 11. The first column (Fig. 11a) shows the rectified left images with a cast shadow. The second column (Fig. 11b) illustrates the results from our proposed approach. The rest of the figure gives the results from SP (Fig. 11c), BP (Fig. 11d), and RT (Fig. 11e) stereo matching algorithms. Those disparity images are all color coded, by which white (255 intensity value) corresponds to the maximum disparity value (130). Since the proposed method relies on shadow information rather than texture, to make a relatively fair comparison, the shadow is kept during the stereo matching procedure. In all the experiments, even though the surfaces do contain texture, the texture is not discriminative enough to establish correspondences. As shown in Fig. 11, all three stereo matching algorithms have difficulty in propagating the correspondences. This is most likely because the low texture surface gives only very sparse feature correspondences, which are not enough to propagate a dense and accurate disparity map. Overall, it is clear that with the addition of shadows, the proposed method significantly outperforms the other three stereo algorithms.

C. 3D Reconstruction Results

The above disparity maps are further processed to get 3D reconstruction results. For the proposed method, the recov-

ered 3D surfaces with and without texture are given in Fig. 12. Those surfaces are displayed using MeshLab, and the snapshots are shown in the figure. Those images in Fig. 12 show that the proposed method is able to recover the 3D surface to a certain degree. For example, in the intestine surface, the deep slopes are nicely recovered. However, it should be noted that errors do occur. For instance, in both the plane and heart examples, the specular reflections cause holes. In the *ex vivo* experiments, the markers themselves have a certain size, thus making the shadow extraction inaccurate when the shadow goes across the markers. In addition, stripes can also be observed in Fig. 12 and they can be reduced if more shadow images are processed. As comparison, 3D reconstruction results of other three methods are also provided in Fig. 13. We recommend to zoom in the figure to have a better understanding of the reconstruction results. Comparing Fig. 12 with Fig. 13, it is clear that the proposed method has great advantage in both accuracy and coverage.

D. Numerical Comparison

To get the ground truth point correspondences for quantitative error analysis, markers are put on the surface and later selected manually from the images, as shown in Fig. 10. Those marker points, (P_l, P_r) , selected from left and right images, serve as ground truth point correspondences. For each point P_l in the left image, P'_r is the calculated corresponding point in the right image. One example of P_r and P'_r in the right image is shown in Fig. 14. The 2D Euclidean distance between P_r and P'_r is named as disparity error and used to reflect the accuracy of disparity maps. In addition, the 3D positions of (P_l, P_r) and (P_l, P'_r) are also computed using triangulation, and the distance between them, named as 3D position error, serves as a measure for the accuracy of the recovered surface, even though the calibration error is inherited in the calculation of the 3D positions. Both disparity error and 3D position error are calculated to compare among the four different methods. Since the disparity maps are sparse and some markers might have no values, for a fair comparison, the nearest valid disparity values (within range [60 130]) are chosen to represent those markers.

The disparity error results of the four methods over the five experiments are given in Table I. The 3D position error comparison results are displayed in Table II. First of all, compare BP with SP and RT, it appears that BP has very low disparity error and 3D position error. In fact, based on our observation, this is most likely because BP method explicitly detects and matches some markers on the image. Both SP and RT do not show such obvious operations. However, even stereo matching methods might get better results because of markers, as show in those tables, the proposed method still significantly outperforms the others. For instance, in the phantom experiments, the disparity errors of the proposed method are within 1.04 pixel, and the 3D position errors of the proposed method are within $0.7mm$. In addition, it is worth to note that in *ex vivo* experiments, both the disparity error and 3D position error are larger than the phantom ones in all four methods. This is probably caused by the wet surface

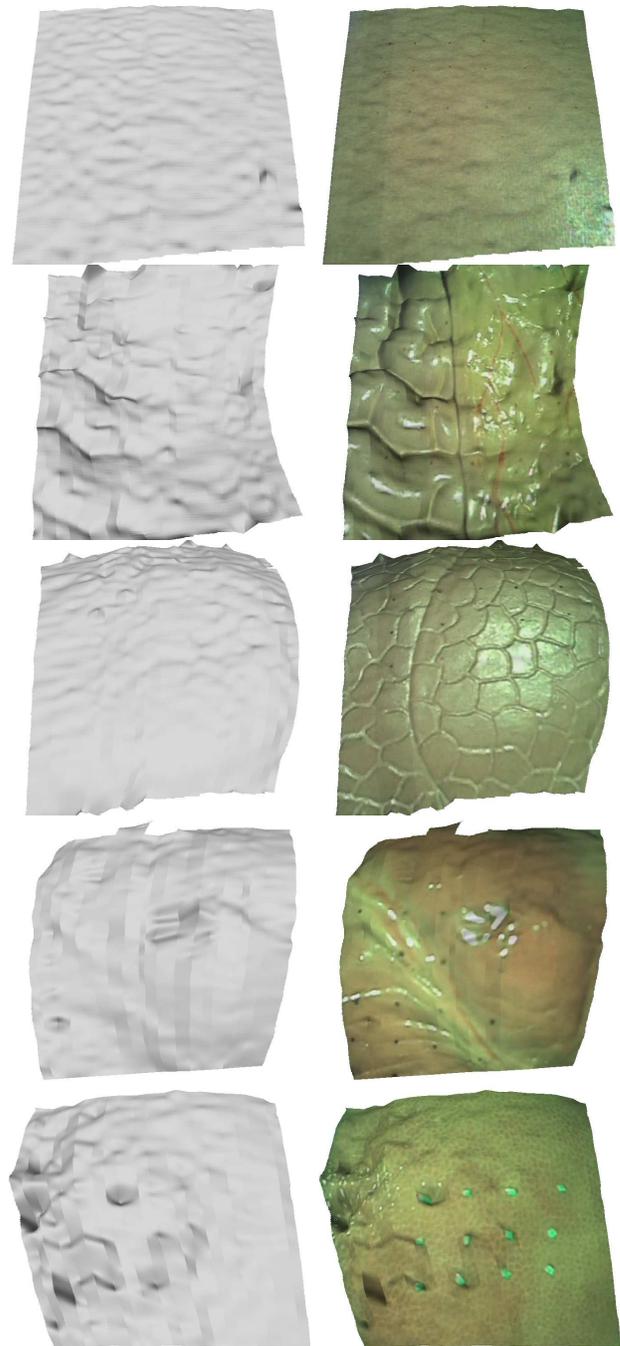


Fig. 12. 3D reconstruction results of flat plane, intestine phantom, lung phantom, heart phantom and porcine liver. The left column shows the recovered 3D model without texture mapping. The right column shows 3D model with texture mapping.

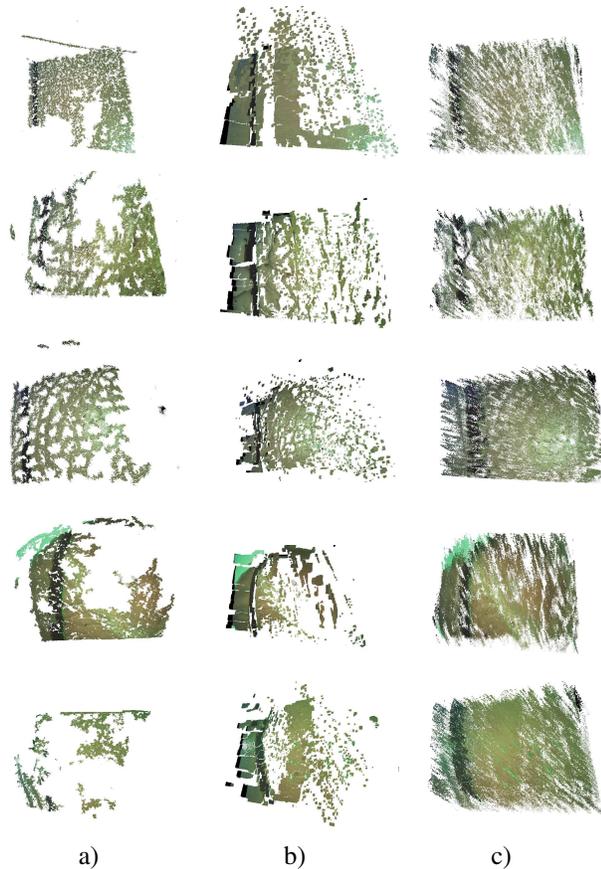


Fig. 13. 3D reconstruction results over five experiments of a) SP method, b) BP method and c) RT method. The first row is experiment on a plane. The second row corresponds to experiment on an intestine phantom. The others are lung, heart and liver respectively.

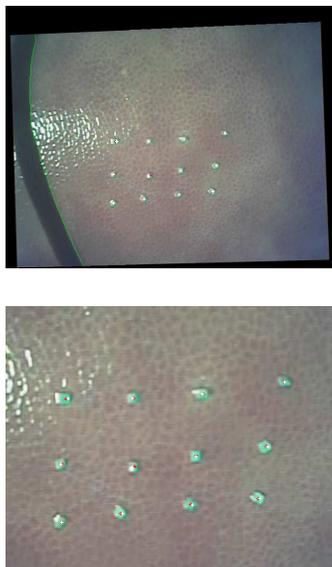


Fig. 14. Illustration of predicted points in right image with ground truth. Red crosses are ground truth points and white crosses denote the calculated points from the disparity map.

TABLE I
DISPARITY ERROR OF THE FOUR METHODS OVER THE FIVE EXPERIMENTS.
ALL THE NUMBER IS IN PIXEL.

Experiments	Proposed method	SP	BP	RT
Plane	0.5387	1.0796	4.0737	31.3568
Intestine	0.5321	1.7709	7.9680	15.1386
Lung	0.9546	2.2666	3.0994	20.2294
Heart	1.0332	9.8639	6.0335	22.3429
Porcine liver	1.3675	28.5919	4.5216	15.4770

TABLE II
3D POSITION ERROR OF THE FOUR METHODS OVER THE FIVE EXPERIMENTS. ALL THE NUMBER IS IN MM.

Experiments	Proposed method	SP	BP	RT
Plane	0.3823	1.7286	5.1175	37.7707
Intestine	0.5923	2.7590	10.7844	24.1257
Lung	0.6553	2.2382	2.5720	27.8008
Heart	0.5834	7.8710	4.4799	21.5661
Porcine liver	1.1406	22.1534	4.2925	11.6948

of the porcine liver, which causes more specular reflections. However, even with the higher complexity in *ex vivo* images, the 3D position error of our method is still within $1.2mm$. On the other hand, in each disparity map, the percentage of pixels whose values are in the range $[60\ 130]$ is recorded in Table III. The numerical comparison of those three tables concludes that our method performs significantly better than the other three both in accuracy and coverage. Next to our method is the SP method, which is followed by BP method. RT method ranks last, probably because it sacrifices the accuracy to achieve real time performance.

E. Robustness Analysis

Since the input of the proposed method comes from shadow curves and calibrated stereo cameras, the accuracy of the final disparity results depends on the precision of shadow extraction and stereo calibration. There are a couple of contributing factors to the shadow border extraction error. The first one is the intensity contrast, that is, a dark shadow and a light background can give better extraction results. Second, the sharpness of the shadow edge directly affects the accuracy of the shadow border. In addition, the synchronization between the two cameras is also an important issue, because only properly-synchronized cameras can guarantee that the left and right shadows correspond to each other. Finally, strong specular reflection can lighten the shadowed area and may disturb the shadow extraction. On the other hand, inaccurate stereo calibration causes error in epipolar lines estimation. Because epipolar lines are used to intersect with the shadow border and establish the point correspondences, the error of

TABLE III
THE PERCENTAGE OF PIXELS IN IMAGE WITH VALID DISPARITY VALUE.

Experiments	Proposed method	SP	BP	RT
Plane	53.87%	30.06%	61.67%	23.34%
Intestine	51.97%	34.00%	60.58%	32.16%
Lung	55.16%	32.58%	65.09%	45.50%
Heart	46.25%	25.25%	53.70%	33.35%
Porcine liver	55.80%	09.74%	46.66%	40.90%

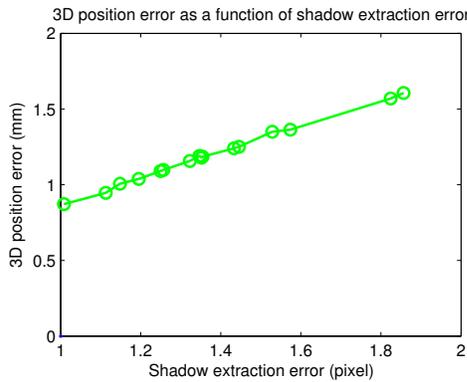


Fig. 15. 3D position error v.s. shadow extraction error.

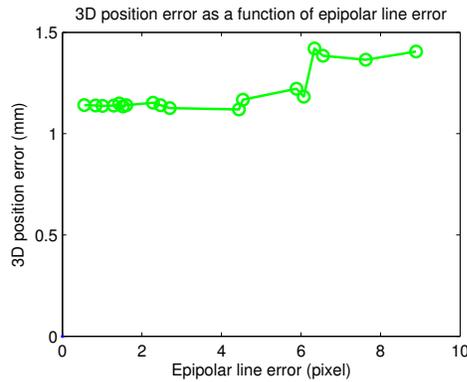


Fig. 16. 3D position error v.s. epipolar line error.

stereo calibration will introduce the horizontal error in point correspondences.

Following the same notation in previous section, we denote $(d_x, d_y) = P'_r - P_r$. For the proposed method, d_x is mainly caused by the shadow extraction error and d_y is the result of the epipolar line calculation error. To analyze how robust the final results are towards the accuracy of shadow border extraction and epipolar lines, uniform pixel noises in different ranges are added to the calculated marker coordinates in the right image, and the corresponding shadow extraction error and 3D position error are recorded. Uniform pixel noises in the range of $[-3, 3]$ are added for shadow border extraction, and uniform noises in the range of $[-10, 10]$ are added for epipolar line computation. In Fig. 15, the 3D position error as a function of the shadow extraction error is displayed, revealing that 3D reconstruction accuracy is linearly proportional to the accuracy of the shadow extraction. This means that the proposed method is robust without significant error propagation towards the accuracy of shadow border extraction. In Fig. 16, it shows that even when the epipolar line error is as large as 10 pixels, the 3D position error is still within 1.5mm . The comparison of Fig. 15 and Fig. 16 empirically indicates that the 3D reconstruction is more sensitive towards the accuracy of the shadow border extraction than the accuracy of epipolar lines.

IV. CONCLUSION

In this paper, we have proposed to use weakly structured light to recover internal organ surfaces for MIS. Typically,

in MIS, the texture of the captured images is not distinctive enough, which makes it difficult to apply the traditional texture-based stereo reconstruction. Instead of relying on texture information, this paper explores the information from shadows generated from surgical tools. The system requirements to use the proposed method are stereo cameras and a separated single-point light source. After shadows are detected in the images, the shadow borders are extracted from both left and right images as shadow curve correspondences. They are later used to intersect with epipolar lines to generate accurate point correspondences for surface reconstruction.

In addition, a novel FSI interpolation is introduced to establish dense correspondences, on which a disparity map is created and 3D surfaces are recovered. The performance of the approach has been validated by phantoms and *ex vivo* images. Disparity maps derived from the proposed method are compared with three popular stereo matching algorithms and demonstrate that the proposed method significantly outperforms the others. Numerical analysis indicates that the accuracy of recovered 3D surface can be up to 0.7mm for phantom models and 1.2mm for *ex vivo* images. Currently, the major holdback from the real time implementation of our method is the shadow generation process, since it takes about 3 seconds to scan through the whole region of interest. In the future, we plan to develop a progressive updating approach that will update the 3D surface model at the areas where dynamic shadows are cast by working with surgical tools during MIS procedures. We are currently collaborating with surgeons for the implementation of our developed systems in MIS and plan to evaluate those systems with *in vivo* experiments in the future work.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful and thoughtful comments and suggestions. They also like to thank for Dr. D. Stoyanov's help with his stereo reconstruction code.

REFERENCES

- [1] D. Stoyanov, M. V. Scarzanella, P. Pratt, and G.-Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in *Proc. MICCAI*, 2010, pp. 275–282.
- [2] J. Totz, P. Mountney, D. Stoyanov, and G.-Z. Yang, "Dense surface reconstruction for enhanced navigation in mis," in *Proc. MICCAI*, 2011, pp. 89–96.
- [3] C.-H. Wu, Y.-N. Sun, and C.-C. Chang, "Three-dimensional modeling from endoscopic video using geometric constraints via feature positioning," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 7, pp. 1199–1211, July 2007.
- [4] 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 *Proc. MICCAI*, 2004, pp. 494–501.
- [5] R. Richa, P. Poignet, and C. Liu, "Efficient 3d tracking for motion compensation in beating heart surgery," in *Proc. MICCAI*, 2008, pp. 684–691.
- [6] R. Richa, A. P. L. Bó, and P. Poignet, "Robust 3d visual tracking for robotic-assisted cardiac interventions," in *Proc. MICCAI*, 2010, pp. 267–274.
- [7] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int J. Comput. Vis.*, vol. 47, pp. 7–42, 2002.
- [8] M. Brown, D. Burschka, and G. Hager, "Advances in computational stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 993–1008, 2003.

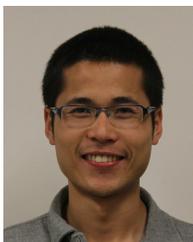
- [9] T. T. Wu and J. Y. Qu, "Optical imaging for medical diagnosis based on active stereo vision and motion tracking," *Opt. Express*, vol. 15, no. 16, pp. 10 421–10 426, Aug 2007. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-16-10421>
- [10] T. T. Wu, T.-H. Cheung, S.-F. Yim, and J. Y. Qu, "Optical imaging of cervical precancerous lesions based on active stereo vision and motion tracking," *Opt. Express*, vol. 16, no. 15, pp. 11 224–11 230, Jul 2008. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-16-15-11224>
- [11] J. D. Ackerman, K. Keller, and H. Fuchs, "Surface reconstruction of abdominal organs using laparoscopic structured light for augmented reality," in *Proc. SPIE*, vol. 4661, no. 1, 2002, pp. 39–46.
- [12] R. K. Mishra, G. B. Hanna, S. I. Brown, and A. Cuschieri, "Optimum shadow-casting illumination for endoscopic task performance," *Arch. Surg.*, vol. 139, no. 8, pp. 889–892, 2004.
- [13] M. Nicolaou, A. James, B. P. L. Lo, A. Darzi, and G.-Z. Yang, "Invisible shadow for navigation and planning in minimal invasive surgery," in *Proc. MICCAI*, 2005, pp. 25–32.
- [14] W. Kunert, T. Kees, H. Raestrup, and G. F. Buess, "The shadow telescope vs. 3-d video system - evaluation in standardised tasks," in *CARS*, 2001, pp. 1052–1055.
- [15] J.-Y. Bouguet and P. Perona, "3d photography using shadows in dual-space geometry," *Int J. Comput. Vis.*, vol. 35, pp. 129–149, 1999.
- [16] A. Agrawal, Y. Sun, J. Barnwell, and R. Raskar, "Vision-guided robot system for picking objects by casting shadows," *Int. J. Rob. Res.*, vol. 29, no. 2-3, pp. 155–173, 2010.
- [17] J. Davis and X. Chen, "A laser range scanner designed for minimum calibration complexity," in *3-D Digi. Imaging and Modeling*, 2001, pp. 91–98.
- [18] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.
- [19] T. Beier and S. Neely, "Feature-based image metamorphosis," *SIG-GRAPH*, vol. 26, no. 2, pp. 35–42, 1992.
- [20] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Mach. Vis. and App.*, vol. 12, pp. 16–22, 2000.
- [21] R. Mishra, *Textbook of Practical Laparoscopic Surgery*. McGraw-Hill, 2009.
- [22] Y. Sun, A. Anderson, C. Castro, B. Lin, R. Gitlin, S. Ross, and A. Rosemurgy, "Virtually transparent epidermal imagery for laparoscopic single-site surgery," in *EMBC*, 2011, pp. 2107–2110.
- [23] C. Castro, S. Smith, A. Alqassis, T. Ketterl, Y. Sun, S. Ross, A. Rosemurgy, P. Savage, and R. Gitlin, "Marvel: A wireless miniature anchored robotic videoscope for expedited laparoscopy," in *ICRA*, may 2012, pp. 2926–2931.
- [24] J.-Y. Bouguet. Camera calibration toolbox for matlab. [Online]. Available: <http://www.vision.caltech.edu/bouguetj/calibdoc/index.html>
- [25] P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," in *Comp. Vis. Pattern Recog.*, vol. 1, 2004, pp. 261–268.
- [26] D. Demirdjian. Real-time stereo library. [Online]. Available: <http://people.csail.mit.edu/demirdji/download/index.html>



Yu Sun is currently an Assistant Professor in the Department of Computer Science and Engineering at the University of South Florida. He was a Post-doctoral Associate at Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA from Dec. 2007 to May 2008 and a Postdoctoral Associate in the School of Computing at the University of Utah from May 2008 to May 2009. He received his B.S. and M.S. degrees in electrical engineering from Dalian University of Technology, Dalian, China, in 1997 and 2000, respectively, and Ph.D. degree in computer science from the University of Utah, Salt Lake City, in 2007. His research interests include biomedical devices, medical imaging and visualization, human-computer-interaction, and robotics.



Xiaoning Qian received the Ph.D. degree in electrical engineering from Yale University, New Haven, CT, in 2005. Currently, he is an Assistant Professor with the Department of Computer Science and Engineering, University of South Florida, Tampa. He was with the Bioinformatics Training Program, Texas A&M University, College Station. His current research interests include computational biology, genomic signal processing, and biomedical image analysis.



Bingxiong Lin received his B.S. degree in mathematics from Jilin University, Changchun, China, in 2008. He received his M.S. degree in computational mathematics from Dalian University of Technology, Dalian, China, in 2010. He is currently working towards his Ph.D. degree in computer science and engineering from the University of South Florida, Tampa, FL. His research interests include surgical robot vision and medical image analysis.