Dense Soft Tissue 3D Reconstruction Refined with Super-pixel Segmentation for Robotic Abdominal Surgery

Veronica Penza · Jesús Ortiz · Leonardo S. Mattos
Antonello Forgione · Elena De Momi

Abstract Purpose: Single Incision Laparoscopic Surgery (SILS) decreases post-operative infections, but introduces limitations in the surgeon's manoeuverability and in the surgical field of view. This work aims at enhancing intraoperative surgical visualization by exploiting the 3D information about the surgical site. An interactive guidance system is proposed wherein the pose of pre-operative tissue models is updated online. A critical process involves the intraoperative acquisition of tissue surfaces. It can be achieved using stereoscopic imaging and 3D reconstruction techniques. This work contributes to this process by proposing new methods for improved dense 3D reconstruction of soft tissues, which allows a more accurate deformation identification and facilitates the registration process.

Methods: Two methods for soft tissue 3D reconstruction are proposed: *Method 1* follows the traditional approach of the block matching algorithm. *Method 2* performs a non-parametric Modified Census Transform to be more robust to illumination variation. The *Simple Linear Iterative Clustering* (SLIC) super pixel algorithm is exploited for disparity refinement by filling holes in the disparity images.

Results: The methods were validated using two video datasets from the Hamlyn Centre, achieving an accuracy of 2.95 mm and 1.66 mm respectively. A comparison with ground truth data demonstrated the disparity refinement procedure: (i) increases the number of reconstructed points by up to 43%; (ii) does not affect the accuracy of the 3D reconstructions significantly.

Conclusion: Both methods give results that compare favourably with the state-of-the-art methods. The computational time constrains their applicability in real-time, but can be greatly improved by using a GPU implementation.

1 Introduction

In open surgery of the abdomen, the trauma of laparotomy wounds is a source of infection or dehiscence and may increase the likelihood of post-operative chest infection, ileus and immobility [30]. The evolution of new minimally invasive approaches, such as Minimally Invasive Surgery (MIS) and Single Incision Laparoscopic Surgery (SILS), allows to perform the surgery by bringing the surgical instruments and endoscope to the surgical site through few or only one small incision, improving the physiological and immuno responses with respect to open surgery [6], and reducing the trauma, without compromising the surgical quality [27, 31].

However, MIS and SILS introduce some limitations, including: (i) the insertion of the instruments through few access ports and their required length reduces the

V. Penza · J. Ortiz · L. Mattos
Department of Advanced Robotics
Istituto Italiano di Tecnologia
via Morego, 30, 16163 Genova, Italy
E-mail: veronica.penza@iit.it, jesus.ortiz@iit.it,
leonardo.demattos@iit.it

A. Forgione
Ospedale Niguarda Ca' Granda, Milano, Italy
P.zza Dell'Ospedale Maggiore, 3, 20162 Milano, Italy
E-mail: antonello.forgione@ospedaleniguarda.it

V. Penza · E. De Momi
Department of Electronics Information and Bioengineering
Politecnico di Milano
P.zza L. Da Vinci, 32, 20133 Milano, Italy
E-mail: veronica.penza@polimi.it, elena.demomi@polimi.it

surgeon's manoeuverability at the surgical site; (ii) the freedom of movement of the endoscopic cameras inside the patient's body is limited, due to the single endoscopic port access; (iii) the structures of interest, such as blood vessels or cancer areas, cannot be viewed from different points of view, compromising the accuracy and safety of the surgery; and (iv) the surgical field visible to the surgeon is limited by the viewing angle of the endoscopic cameras. Such factors extend the learning curve for surgeons and increase operating times [28].

The recent development of advanced robotic systems for MIS and SILS has been motivated by drawbacks related to the manoeuvrability of the surgeon [28]. Robot-assisted devices can greatly help to restore the intuitiveness of operations in such procedures, but this is largely dependent on system's surgical vision. For example, the da Vinci® Surgical System [8] allows intuitive execution of surgery and hand-eye coordination through 3D vision and the configuration of its surgical console. Thus, a requirement in robotic MIS and SILS is to provide an appropriate surgical vision system optimally coupled to the control of the robotic arms. However, in many cases this many not be enough, especially in SILS operations: once the robot is inside and moving around the patient's body, it is easy for the surgeon to get disoriented since the pose of the robot cannot be observed (or easily inferred) from the outside. This requires an extra learning phase even for experienced laparoscopy surgeons, and also poses a safety risk for the operation.

Computer assisted technologies can enhance the view of the surgical field. Augmented Reality (AR) or Augmented Virtuality (AV) systems can provide a more comfortable and efficient environment for the surgeon during surgery [15]. For example, in [16], pre-operative information about the disease and a surgical plan are fused with the intra-operative visualization of the surgical field.

Nevertheless, aligning multimodal pre-operative patient data is highly challenging in soft-tissue surgery since the anatomy undergoes significant changes between the data acquisition (pre-operative) phase and the surgical procedure (intra-operative) due to different factors [33]: (i) different pose of the patient with respect to the pose in which the pre-operative image was taken; (ii) CO2 abdominal insufflation for increasing the working volume (*pneumoperitoneum*); (iii) instrument tissue interaction; (iv) heart beat; (v) breathing; etc. Overcoming these challenges requires online estimation of the tissue deformations to correct the pre-operative plan, which can be done, for example, through methods of depth estimation from stereo endoscopic images (3D reconstruction). The denser the reconstruc-

tion, the more accurate the deformation identification will be. The real-time 3D information from the surgical site can also help the surgeon to understand the relationship between the organs and the robotic devices. Moreover, for robotic SILS procedures the surgical site visualization issues can be overcome by introducing a virtual environment including a dynamic model of the robot within a surgical scene recreated from pre-operative information. This virtual environment can be updated in real-time by exploiting 3D reconstruction of the soft tissues [4, 22].

In this paper, we present two methods for dense 3D reconstruction of soft tissue from endoscopic images, exploiting a super pixels approach for the disparity map refinement. The dense surface reconstruction algorithm facilitates the registration process between the pre-operative model and the intra-operative site. The methods form the basis of the *virtual assistive system*, introduced in [20], for intra-operative guidance and improved surgical safety for abdominal SILS procedures (Fig. 1). The system provides: (i) surgical guidance by exploiting patient-specific pre-operative 3D models obtained from CT scans to improve the operative vision; and (ii) improved surgical safety by providing the visualization of the pose of the robotic devices.

The paper is organized as follows: After a brief review on the state-of-the-art in Section 2, a description of the *virtual assistive system* and the contribution of this paper are presented in Section 3. The complete algorithms for soft tissue surface reconstruction are described in Section 4. Section 5 presents the evaluation of the proposed methods, and the obtained results are presented and discussed in Section 6 and Section 7 respectively.

2 State of the art

The surface reconstruction of man-made environments using stereo images is a well understood concept [14, 23]. However, 3D surface reconstruction of surgical endoscopic images is still an active area of research due to some challenging aspects that include: (i) the application in surgery requiring high accuracy and robustness in order to ensure patient safety; (ii) difficulties arising due to large lens distortion, many texture-less areas, occlusions introduced by the surgical tools, specular highlights, smoke and blood produced during the intervention [25]; etc. Many approaches have been proposed in literature in order to achieve more reliable, robust, and real-time depth measurement to represent a deforming environment. Some approaches are based on feature detection and tracking [13], allowing the recovery of a sparse surface. In [26], a semi-dense reconstructed

surface is obtained using feature detection and propagating the information into neighboring regions. The robustness of these methods depends on the availability of stable features. Other solutions are based on finding correspondences in the pixel intensities to give dense reconstructed surfaces. In [22], dense GPU-enhanced 3D reconstruction is obtained using the Hybrid Recursive Matching algorithm performing a non-parametric transformation on the images. A non-parametric image transformation with model based methods is also exploited in [24] for dense 3D reconstruction, with the assumption of a smooth and continuous surgical surface. [19] presents a comparative study of such state-of-the-art methods. Recently, color-based segmentation was used in stereo methods for depth estimation [11]. Such techniques first segment a reference image into regions and then label each region with disparity values, preserving the boundaries of the objects. For example, the segmentation is used in combination with belief propagation within a Markov Random Field framework in [32]. However, segmentation algorithms are computationally expensive and infrequently used on surgical images.

3 Research Overview

The overall goal of our research is to enhance the visual information provided to the surgeon during robotic SILS procedures in the abdomen. Towards this end, a first prototype of a *virtual assistive system* was presented in [20]. An overview is shown in Fig. 1. In the pre-operative phase, the surgeon extracts 3D surface models from a CT scan dataset of the patient. These models are inserted into a virtual environment. A stereo camera calibration and a real-virtual camera calibration are performed. During the surgery, a semi-automatic registration of the virtual abdomen on the real patient creates a connection between the reality and the virtual environment, which consists of:

1. View of the entire abdomen model, where it is possible to visualize at run-time the motion of the robot and the surgical tools. The possibility to change the transparency of the model of the skin allows the surgeon to plan the entry point, adjusting the access region with respect to the target to be reached.
2. A virtual camera image plane, which shows the structures from the same point of view as the real endoscopic camera. This can enable the visualization of the hidden structures using the transparencies function. A zoom feature allows the surgeon a wider field of view if desired.
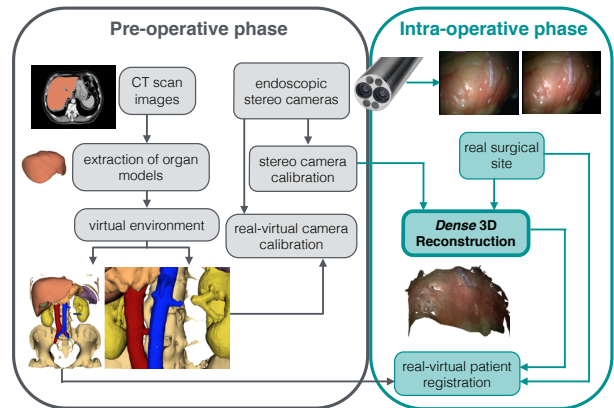

Fig. 1 *Virtual Assistive System.* The left side shows the pre-operative phase with the processing of the CT scan images and the cameras calibration. The right side shows the intra-operative phase with the processing of the surgical images and the integration between the real and virtual images. The endoscope shown is from the Da Vinci® system

The current design of the *virtual assistive system* from [20] is reliable only under the assumption that the patient remains in the same condition as he/she was during the CT scan phase. However, this is an ideal condition, and as noted earlier, the changes in anatomy from pre-operative phase to intra-operative phase are unavoidable. To overcome this drawback, the 3D reconstruction of the tissues at run-time can serve as the keystone for updating the pre-operative tissue models. Moreover, by knowing the pose of the tissue surface with respect to the robotic devices, it would be possible to define areas to be protected, e.g., main vessels or vital structures, improving the safety of the surgery.

3.1 Research Contributions

Two techniques for 3D surface reconstruction of soft tissue are investigated in this paper:

1. *Method 1* follows the traditional approach of the block matching algorithm [23].
2. *Method 2* exploits a non-parametric transformation to make the stereo matching more robust to illumination variations.

The paper introduces a novel method to enhance the density of the reconstructed surface from these methods through disparity refinement based on Simple Linear Iterative Clustering (SLIC) super-pixels algorithm [1].

The algorithms introduced here are designed to enhance the visualization during robotic SILS procedures and are integrated into the *virtual assistive system* as highlighted in Fig. 1.

4 Materials and Methods

The workflow for the soft tissue 3D reconstruction algorithm using the two methods is shown in Fig. 2 and it is described in detail in the following subsections.

4.1 Pre-Processing

Surgical images acquired from an endoscope need pre-processing in order to facilitate the search for pixel correspondences. For greater accuracy in the 3D reconstruction we calibrate the cameras using the OpenCV Library [17], obtaining the intrinsic and extrinsic parameters. The images are then undistorted and rectified, so that the corresponding pixels lie on the same horizontal line (epipolar line) [9]. Generally, the stereo cameras are on the tip of the inserted endoscope device and in proximity to a light source. The two cameras are exposed differently to the light source due to their different position with respect to it. This can produce over or under exposure and differences in brightness/exposure between the images. Image processing removes such artifacts in the images and enhances the identification of the correspondences in texture-less and homogeneous areas. These steps are described in the following paragraphs.

Specularity Removal. Clinical images are usually affected by specular reflections due to the tissue characteristics and the proximity of the light source. The specular reflections appear as bright regions in the endoscopic images and can be mistaken for regular/irregular tissue. The algorithm described in [3] is implemented for removing specular highlights.

Image Equalization. Histogram equalization is performed to improve contrast and enhance details that are over or under-exposed. The method used is a function from the OpenCV Library (*equalizedHist*). The equalization is performed only in *Method 1*, since the Census transform in *Method 2* already makes it robust to illumination variation.

4.2 Surface Reconstruction

The two cameras of the endoscope record the surgical field from two different points of view. The difference in position of an object between the left and right images is called *disparity* (d). Knowing the distance between the cameras (s) and the focal length of the cameras (f), the depth, i.e. the z coordinate of the reconstructed point, can be calculated with the following equation:
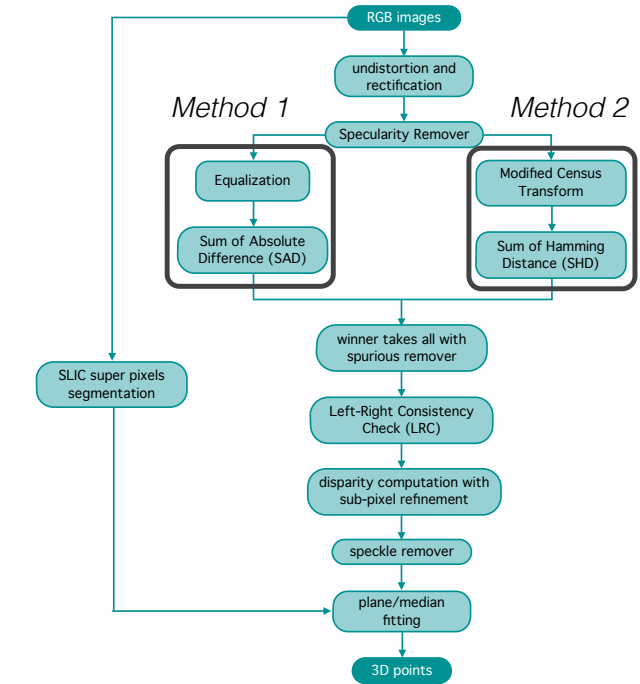

Fig. 2 Workflow of the proposed surface extraction algorithm. Two methods are proposed: *Method 1* one based on SAD and *Method 2* based on Census transform. Both methods are improved using the disparity refinement with superpixel segmentation

$$depth(i,j) = \frac{(f \cdot s)}{disparity(i,j)} \qquad (1)$$

where i and j are respectively rows and columns of the image.

Stereo correspondence is the process of identification of the matching between each pixel of the left and the right images (im_l and im_r). The process can be divided in three steps, as stated in [23]: (i) Matching cost computation; (ii) Aggregation cost computation; and (iii) Disparity computation. The two methods are described below:

Method 1

1. **Matching cost computation:** This method uses the *absolute intensity difference* (AD) [10] operation to calculate the similarity between two pixels.
2. **Aggregation cost computation:** The aggregation of the matching cost is done by summing the AD of a window of size n x m, i.e., by computing the Sum of Absolute Differences (SAD). Eq. 2 shows the SAD value for a pixel (i,j):
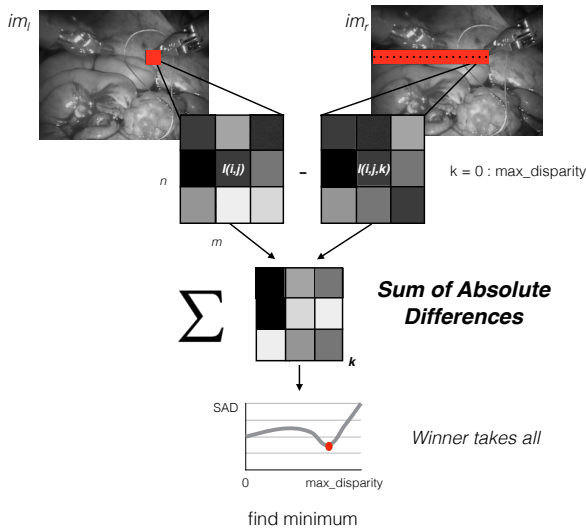
Fig. 3 Disparity computation algorithm for *Method 1*. It uses the intensity of the pixels and the Sum of Absolute Differences (SAD) operation for the aggregation

$$SAD(i,j,d) =$$
$$\sum_{\substack{h=-m/2 \\ h=-n/2}}^{\substack{m/2 \\ n/2}} |I_l(i+h,j+k) - I_r(i+h-d,j+k)| \quad (2)$$

where I_l and I_r is the intensity of each pixel of im_l and im_r respectively, and d is the disparity. The matching cost computation and aggregation cost of *Method 1* are shown in Fig. 3.

3. **Disparity computation:** The computation of the final disparity involves choosing the disparity at each pixel that is associated with the minimum of the aggregation cost value. The *Winner Takes All* (WTA) strategy from [23] is used to find the minimum aggregation cost value at each pixel. Two optimizations are used to reduce the number of incorrect values (from an incorrect matching). With the first optimization, the disparity is defined to be invalid if two minimum aggregation cost values in the same image are within a threshold [18]. This operation helps to invalidate incorrect pixels on uniform surfaces. The second optimization, a Left-Right Consistency (LRC) Check, is performed in order to invalidate half-occluded pixels, i.e., objects viewed in one image but not in the other.

Method 2

1. **Matching cost computation:** The non-parametric Census transform [2] is applied on im_l and im_r, with

a sparse modified approach from [12]. It converts each pixel inside a moving window into a string of bits $C(i,j)$ (see Eq. 3), representing neighbor pixels according to a comparison with the central pixel and the mean value of the pixels inside the window (Eq. 4).

$$C(i,j) = \bigotimes_{h=-m/2}^{m/2} \bigotimes_{k=-n/2}^{n/2} \xi(I(i,j), I((i,j) + (h,k))) \quad (3)$$

where $\bigotimes$ denotes the concatenation to a bit string. ξ is the function for comparing the two intensities, defined as:

$$\xi(I,I') = \begin{cases} 00 & I' \leq min(I,\bar{I}) \\ 01 & I' < I \\ 10 & I' > I \\ 11 & I' \geq max(I,\bar{I}) \end{cases} \quad (4)$$

where I' is the moving pixel in the window compared with the central pixel I and $\bar{I}$ is the average of the pixel intensities inside the window.

The similarity between the two Census transformed images is analysed using the *Hamming Distance* (HD) measure applied to a moving window. HD compares bit strings representing the pixels and identifies the number of positions at which the corresponding bits are different, as shown in Eq. 5.

$$HD(i,j,d) = \sum_{b=0}^{m \cdot n} C_l(i,j)_b \oplus C_r(i-d,j)_b \quad (5)$$

where C_l and C_r are the Census transform of the left and right images respectively, $\oplus$ is the XOR operator and b represents the bit string.

The HD is computed only for a chessboard pattern of pixels inside the window in order to decrease the computational time of the algorithm.

2. **Aggregation cost computation:** The aggregation of the matching cost is done by summing the Hamming Distance over the window of size n x m, computing the Sum of Hamming Distance (SHD). The matching cost computation and aggregation cost of *Method 2* is shown in Fig. 4.

3. **Disparity computation:** This step uses the same strategy as that for the *Method 1*.

To avoid separated layers in the reconstructed surface resulting from pixel-level precision, a sub-pixel refinement is applied using a parabola fitting [29] in both methods. A speckle removal filter is also applied in order
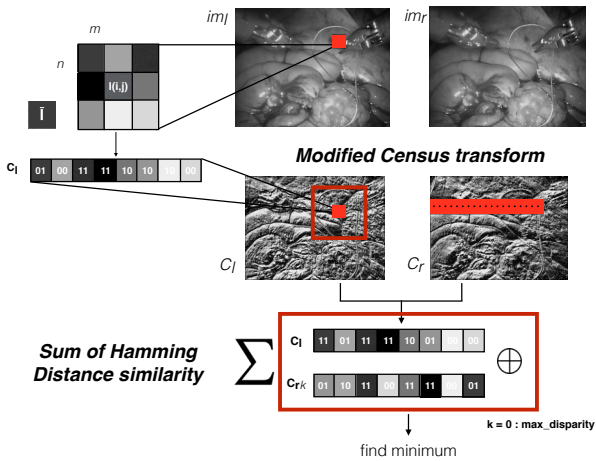
Fig. 4 Disparity computation algorithm for *Method 2*. It uses Sparse Modified Census transform and Hamming distance as cost and aggregation function. The disparity value is the one associated with the minimum of the aggregation cost value

to remove small artefacts, i.e., regions of large and small disparities that can be generated near the boundaries of shapes. This method is adopted from the OpenCV Library [17] function (*filterSpeckles*).

4.3 Disparity Refinement

The disparity map, computed so far from the procedures described above, can contain invalid values (holes), which can compromise the usefulness of the reconstructed surface. As mentioned in Section 3, the 3D reconstruction of the surgical field serves to update the pose of the pre-operative models considering the real organ deformations, and also to update the pose of safety areas to be protected from the surgical tools during surgery. Problems can arise if the motion and the deformation of the soft tissue are not identified, i.e., if the reconstructed surgical field is not dense and accurate.

To address these issues, a method to exploit the over-segmentation of the reference image Im_l is introduced here, which further refines the results and fills the holes in the disparity map. Im_l is segmented using the Simple Linear Iterative Clustering (SLIC) super-pixels algorithm [1], which adapts the k-means clustering approach to generate super-pixels efficiently and in less computational time. At each super-pixel, a label L is assigned. Then $Im_l(L)$ is set as a group of pixels from the left image with the same label, i.e., belonging to the same super-pixel. The same labels are assigned to the disparity map, $disparity(L)$. Each super-pixel is ideally a homogeneous area with similar or at least continuous depth. This information can be exploited in order to fill
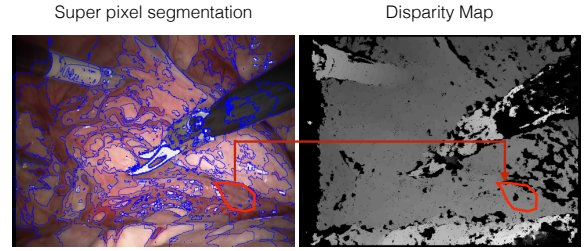

Fig. 5 The superpixel areas (on the left) can be used to fill gaps in the disparity image (on the right)

the holes in the disparity map, avoiding large discontinuities within a super-pixel, as shown in Fig. 5. We apply two different strategies depending on the valid values inside a label:

1. **Plane fitting:** The invalid disparity values are fitted to a plane which is computed on the valid values of the disparity map. The Locally Optimized RANSAC (LO-RANSAC) method [5] is used to find the optimal plane parameters in the pixel space, using the model of the plane described in Eq. 6.

$$disparity(i, j) = a \cdot i + b \cdot j + c \qquad (6)$$

   where a, b and c are the coefficients of the fitted plane.
2. **Median fitting:** If the plane fitted on the valid values is not reliable (according to the LO-RANSAC results), the invalid disparity values are replaced by the median of the valid disparity values.

5 Evaluation

5.1 Quantitative Evaluation

This section presents the evaluation of the methods for the 3D surface reconstruction. The performances of *Method 1* and *Method 2* are evaluated against the metrics of accuracy, computational time, and percentage of matched pixels.

The Hamlyn Centre Laparoscopic/Endoscopic Video dataset [7], with an associated ground truth representing a surgical scene, is used for the evaluation.

Stereo images of a silicon heart phantom were used in order to validate the results and ensure the proposed methods are robust and reliable when applied to a surgical scenario. The dataset consisted of intrinsic, extrinsic, and distortion parameters for the cameras and a ground truth from a CT scan. We used 20 frames from two videos, *heart1* and *heart2*. These frames are

named $h1_i$ and $h2_i$ respectively, where i indicates the frame number. The ground truth sets are called $g1_i$ and $g2_i$. The point clouds computed using the two methods are $pc1_i$ and $pc2_i$. The use of this dataset for the evaluation of the proposed system was preferred over other standard datasets, such as the Middlebury evaluation [22], because these provide only static images and do not show surgical scenarios, which present specific issues related to endoscopic imaging.

The error evaluation consists of computing the Euclidean distance between the corresponding points from the extracted point cloud and the rectified ground truth, as in Eq. 7. The final *accuracy* of each point cloud is computed as the median value of these Euclidean distances. *Method 1* and *Method 2* are evaluated on the two datasets *heart1* and *heart2* with and without the application of the disparity refinement exploiting SLIC Segmentation (hereinafter called SLIC refinement).

$$error(i, j) =$$
$$\sqrt{(pc_{x_{i,j}} - g_{x_{i,j}})^2 + (pc_{y_{i,j}} - g_{y_{i,j}})^2 + (pc_{z_{i,j}} - g_{z_{i,j}})^2}$$
(7)

where $pc_{x_{i,j}}$, $pc_{y_{i,j}}$ and $pc_{z_{i,j}}$ are the 3D coordinates of a point from the computed point cloud, and $g_{x_{i,j}}$, $g_{y_{i,j}}$ and $g_{z_{i,j}}$ are the coordinates from the ground truth. Since the point cloud and the ground truth are represented as 2D maps, we can simply calculate this error for each point considered as a pixel of the image.

A non-parametric test (Wilcoxon $p < 0.05$) was performed on the point clouds to test if the accuracy before and after SLIC refinement was significantly different, for both *Method 1* and *Method 2*. The same test was also performed to verify if the accuracy obtained with *Method 1* and *Method 2* was significantly different.

The *percentage of matched pixels* is computed as the ratio of the valid reconstructed points identified in the region of interest with respect to the total number of points of the image. This region takes into consideration only the area of the image with a pixel intensity > 32, eliminating the areas where the heart phantom is not visible. An example of the region of interest used for the evaluation is shown in Fig 6.

The *computational time* is the time of execution of the algorithms. The code was developed in C++, using OpenCV for the management of the images with Robot Operating System (ROS) as framework [21]. The program was running on a system with GNU/Linux operating system, and a CPU Intel Core i7-4820K with 4 cores and hyper-threading (8 virtual cores).

The parameters and labels used during the evaluation are listed in Table 1. The window dimension for the
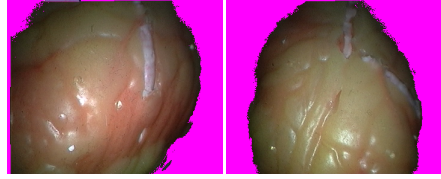

Fig. 6 Example of the region of interest for *heart1* and *heart2* models. The points marked with magenta color were not included in the evaluation of the methods

Census transform and the Aggregation cost of *Method 2* are taken from [12], where the authors identify the optimal parameters taking into consideration accuracy and computational time.

Table 1 Parameters used in the evaluation

Parameters	Value
Census window	9x9
Census Block Size	11x11
SAD Block Size	11x11
Threshold Spurious remover	10
Threshold LRConsistency Check	4
LO-RANSAC max iteration	100
Number of sample frames	20
Labels	Value
Method 1	M1
Method 1 with SLIC refinement	M1R
Method 2	M2
Method 2 with SLIC refinement	M2R

5.2 Qualitative Evaluation

The target application for this research is robotic SILS procedure for the abdomen. Unfortunately, video datasets with ground truth information are not available for this anatomical region. Therefore, qualitative evaluations of the proposed methods were performed using the heart data set from the Hamlyn Center. This data set approximates the envisioned application area of our system well, including motions of structures/organs and the presence of surgical instruments in the field of view.

6 Results

Fig. 7 shows a sample image from *heart1* processed with both methods, following the workflow of Fig. 2. The figure demonstrates the presence of the invalid values in the disparity map as well as the almost complete correction of the map with the application of SLIC refinement. The effect is clearly seen considering the percentage of matched pixels in Fig. 8. The application of
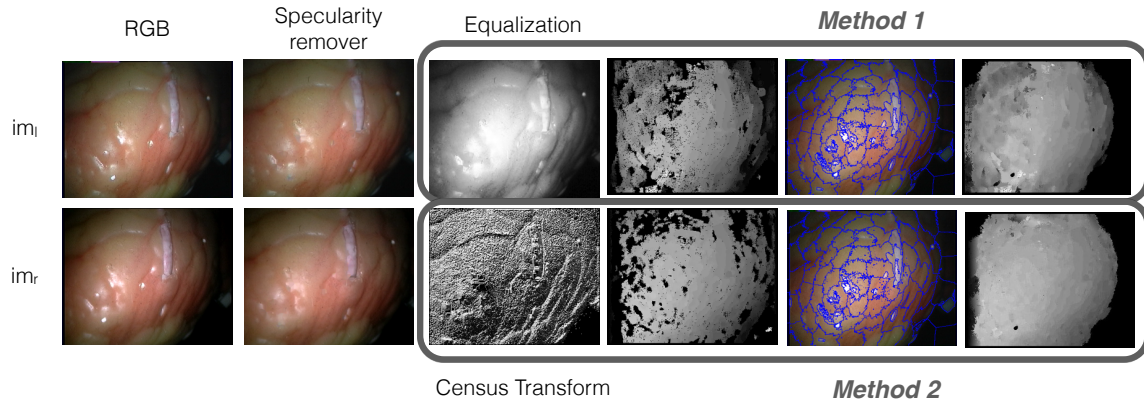
Fig. 7 Example of the application of the workflow presented in this paper on one frame of *heart1* dataset. From left to right, the RGB stereo images I_l and I_r, the stereo images after the specularity removal filter and *Method 1* at the top and *Method 2* at the bottom. The *Method 1* images show the equalization of the image, the disparity map calculation, the SLIC segmentation and the successive refinement of the disparity map. The *Method 2* images show the Census transform, the disparity map calculation, the SLIC segmentation and the successive refinement of the disparity map

the SLIC refinement increases the percentage of valid values to more than 70% in this case.

A box plot representing the *accuracy* is shown in Fig. 9. These results show an accuracy of $3.17mm$ and $3.27mm$ using M1 and M1R and an accuracy of $1.68mm$ and $1.77mm$ for M2 and M2R respectively. We can also see how the variance of these results is similar, indicating good consistency between the presented methods.

Fig. 10 shows a colormap representing the accuracy of the reconstructed surface computed with the methods. Here again, the result of applying the SLIC refinement is demonstrated and how the filled holes present a low error with respect to the ground truth.

For the impact of the SLIC refinement on the accuracy, the non-parametric test did not reject the null hypothesis for both methods. This implies that the SLIC refinement did not have a statistically significant impact on the accuracies of M1 and M2. The same test rejected the null hypothesis in the comparison between the values obtained with M1 and M2. This implies that M2 is statistically better and provides significantly improved accuracy over M1.

We can see the results of the evaluation in Table 2, which shows the median values for the *accuracy, percentage of matched pixels*, and *computational time*. Both methods have a similar computational time, around $1.21s$ per frame (with an image resolution of 288x360), and the SLIC refinement adds an overhead of only $0.1s$ in both cases.

The results from the qualitative evaluation are shown in Fig. 11. We can see that applying the methods on an image representing a more complex surgical scenario, with different abdominal structures and instruments visible in the field of view, the methods are still able

to reconstruct the surface, and SLIC refinement seems to fill correctly the invalid values of the disparity map.

Table 2 Median values of the *accuracy, percentage of matched pixels*, and *computational time* for both methods and data sets, with and without SLIC refinement

	M1	M1R	M2	M2R
Accuracy [mm]				
heart 1	2.95	3.11	1.66	1.75
heart 2	3.38	3.43	1.70	1.79
Percentage of matched pixels [%]				
heart 1	57.5	72.6	51.9	72.6
heart 2	55.4	66.5	44.7	66.5
Computational time [s]				
heart 1	1.21	1.30	1.21	1.29
heart 2	1.21	1.31	1.21	1.30

7 Discussion and Conclusion

The research presented in this paper aims at improving a dense 3D surface reconstruction of surgical environments, specifically to enable enhanced visualization during robotic SILS procedures of the abdomen. The 3D surface information from the surgical scene can be exploited to update the pose of pre-operative models extracted from a CT scan. These models, inserted into a virtual assistive system, can provide guidance information to the surgeon, helping to improve the quality and safety of procedures through augmented visualization. In addition, information about the pose of the surgical robot with respect to tissue surfaces can assist in defining and enforcing safety constraints intra-operatively.
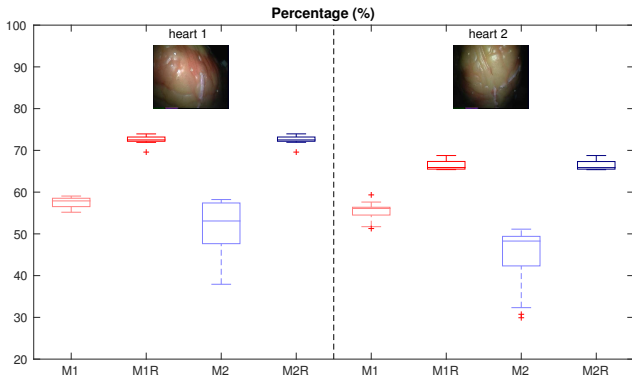
Fig. 8 Box plot representation of the percentage of matching pixels (pixels that are considered valid) with *Method 1* and *Method 2* evaluated on *heart1* and *heart2* without and with the SLIC refinement
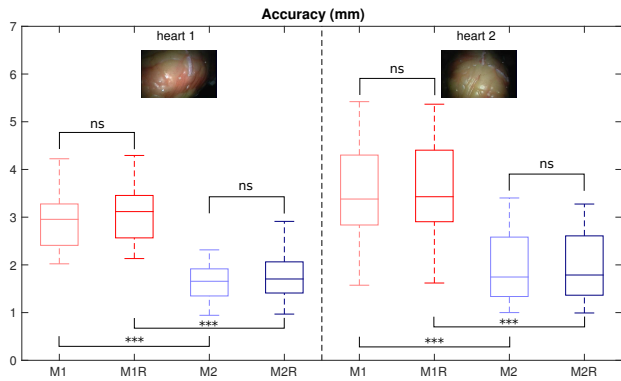

Fig. 9 Box plot representation of the accuracy obtained with *Method 1* and *Method 2* evaluated on *heart1* and *heart2* without and with the SLIC refinement
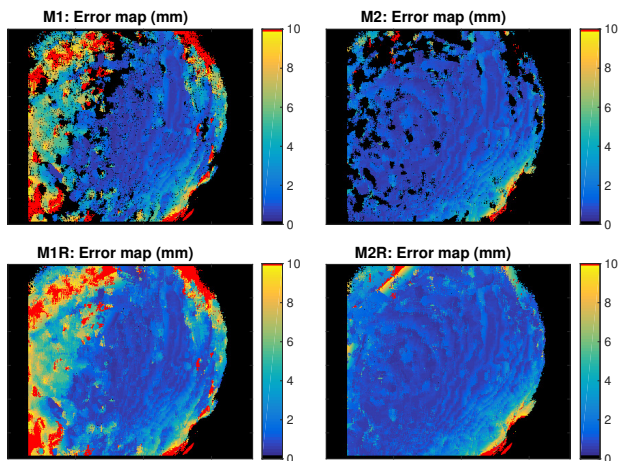

Fig. 10 2D representation of the reconstructed 3D surface, obtained from *Method 1* and *Method 2* with and without SLIC refinement. The color represents the error
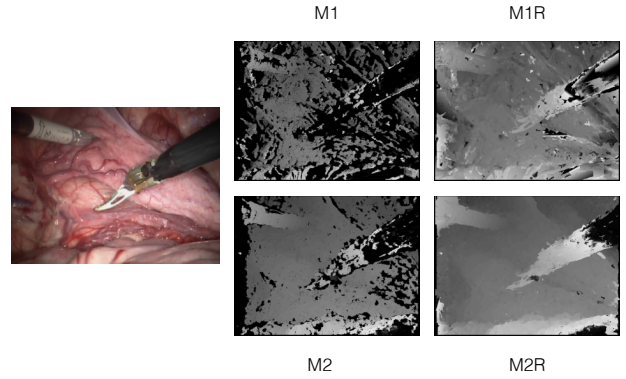

Fig. 11 Qualitative results obtained by the application of M1 and M2 on a surgical video. The image on the left shows Im_l. The images on the right show the disparity maps obtained with the different methods presented in this paper

Two methods for 3D surface reconstruction were presented based on a traditional approach (M1) and a non-parametric transformation approach (M2). The comparison between the two methods shows that M2 has a significantly higher accuracy respect to M1 (around $1.7mm$). According to our team of surgeons, for the proposed application, the overall system accuracy should be below $10mm$. This is motivated by the fact that abdominal surgery cannot be considered as a precision surgery and it is characterized by the presence of motions and deformations of the organs. For this reason, the obtained accuracy with both methods can be considered as a good result since it should allow to comply with the required accuracy.

Regarding the valid reconstructed points, M1 provides a higher percentage of matching pixels. The methods were also complemented to enhance the density of the reconstruction using the SLIC disparity refinement. With the refinement, the percentage of matching pixels increases in both methods, reaching the 72.6% in the best case. The refinement provided an improvement in the matching percentage up to 23% in M1 and 43% in M2, without any statistically significant impact on the accuracy.

A comparison of our error results with those from [19], which tests different state-of-the-art algorithm using the same video dataset, demonstrated that our errors are:

- lower as compared to Stereo Block Matching (SBM) and variational (Svar) algorithms.
- similar magnitude as compared to Stereo Semi-Global Matching (SSGBM).

The strength of the methods presented in this paper is the usage of the SLIC Super Pixel algorithm to obtain high density valid disparity map, which can be exploited for augmented reality applications.

In comparison to [22] the accuracy achieved is slightly worse (mean error $2.47mm$ against $1.55mm$), while the percentage of pixel reconstructed obtained exploiting the SLIC refinement is higher (72.6% against 66.2%).

In the extension of this research, future work will be focused on the development of a real-time implementation of the proposed algorithm, potentially based on a hybrid CPU-GPU processing framework. The expected speed-up is between 10 and 30 times for the current resolution of the images, which would allow reaching the target framerate of $25fps$. For higher resolution images, in order to maintain a real-time execution of the algorithm, a sub-sampling could be applied. The motion tracking of the organs will also be investigated in combination with the soft tissue surface reconstruction. Finally, the components will be integrated into the *virtual assistive system*.

References

1. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S (2012) Slic superpixels compared to state-of-the-art superpixel methods. Pattern Analysis and Machine Intelligence, IEEE Transactions on 34(11):2274–2282
2. Banks J, Bennamoun M, Corke P (1997) Nonparametric techniques for fast and robust stereo matching. In: TENCON'97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications., Proceedings of IEEE, IEEE, vol 1, pp 365–368
3. Barbalata C, Mattos L (2014) Laryngeal tumor detection and classification in endoscopic video. IEEE journal of biomedical and health informatics
4. Bergen T, Wittenberg T (2014) Stitching and surface reconstruction from endoscopic image sequences: A review of applications and methods
5. Chum O, Matas J, Kittler J (2003) Locally optimized ransac. In: Pattern Recognition, Springer, pp 236–243
6. Ee WWG, Lau WLJ, Yeo W, Von Bing Y, Yue WM (2014) Does minimally invasive surgery have a lower risk of surgical site infections compared with open spinal surgery? Clinical Orthopaedics and Related Research® 472(6):1718–1724
7. Giannarou S, Visentini-Scarzanella M, Yang GZ (2013) Probabilistic tracking of affine-invariant anisotropic regions. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35(1):130–143, DOI 10.1109/TPAMI.2012.81
8. Guthart G, Salisbury Jr JK (2000) The intuitivetm telesurgery system: Overview and application. In: ICRA, pp 618–621
9. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge university press
10. Kanade T, Okutomi M (1994) A stereo matching algorithm with an adaptive window: Theory and experiment. Pattern Analysis and Machine Intelligence, IEEE Transactions on 16(9):920–932
11. Klaus A, Sormann M, Karner K (2006) Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, IEEE, vol 3, pp 15–18
12. Ma L, Li J, Ma J, Zhang H (2013) A modified census transform based on the neighborhood information for stereo matching algorithm. In: Image and Graphics (ICIG), 2013 Seventh International Conference on, IEEE, pp 533–538
13. Mountney P, Stoyanov D, Yang GZ (2010) Three-dimensional tissue deformation recovery and tracking. Signal Processing Magazine, IEEE 27(4):14–24
14. Musialski P, Wonka P, Aliaga DG, Wimmer M, Gool L, Purgathofer W (2013) A survey of urban reconstruction. In: Computer Graphics Forum, Wiley Online Library, vol 32, pp 146–177
15. Nicolau S, Goffin L, Soler L (2005) A low cost and accurate guidance system for laparoscopic surgery: Validation on an abdominal phantom. In: Proceedings of the ACM symposium on Virtual reality software and technology, ACM, pp 124–133
16. Nicolau S, Soler L, Mutter D, Marescaux J (2011) Augmented reality in laparoscopic surgical oncology. Surgical oncology 20(3):189–201
17. OpenCV (2015) URL http://opencv.org/
18. Ortiz J, Calderon H, Fontaine J (2010) Disparity map computation on scalable computing. In: Mechatronics and Embedded Systems and Applications (MESA), 2010 IEEE/ASME International Conference on, IEEE, pp 301–306
19. Parchami M, Cadeddu JA, Mariottini GL (2014) Endoscopic stereo reconstruction: A comparative study. In: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, IEEE, pp 2440–2443
20. Penza V, Ortiz J, De Momi E, Forgione A, Mattos L (2014) Virtual assistive system for robotic single in-

cision laparoscopic surgery. In: 4th Joint Workshop on Computer/Robot Assisted Surgery, pp 52–55

21. Quigley M, Conley K, Gerkey BP, Faust J, Foote T, Leibs J, Wheeler R, Ng AY (2009) Ros: an open-source robot operating system. In: ICRA Workshop on Open Source Software

22. Röhl S, Bodenstedt S, Suwelack S, Kenngott H, Müller-Stich BP, Dillmann R, Speidel S (2012) Dense gpu-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration. Medical physics 39(3):1632–1645

23. Scharstein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision 47(1-3):7–42

24. Schoob A, Kundrat D, Kleingrothe L, Kahrs LA, Andreff N, Ortmaier T (2014) Tissue surface information for intraoperative incision planning and focus adjustment in laser surgery. International journal of computer assisted radiology and surgery pp 1–11

25. Stoyanov D (2012) Surgical vision. Annals of biomedical engineering 40(2):332–345

26. Stoyanov D, Scarzanella MV, Pratt P, Yang GZ (2010) Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010, Springer, pp 275–282

27. Tan HJ, Wolf JS, Ye Z, Hafez KS, Miller DC (2014) Population level assessment of hospital based outcomes following laparoscopic versus open partial nephrectomy during the adoption of minimally invasive surgery. The Journal of urology 191(5):1231–1237

28. Taylor G, Barrie J, Hood A, Culmer P, Neville A, Jayne D (2013) Surgical innovations: Addressing the technology gaps in minimally invasive surgery. Trends in Anaesthesia and Critical Care 3(2):56–61

29. Tian Q, Huhns MN (1986) Algorithms for subpixel registration. Computer Vision, Graphics, and Image Processing 35(2):220–233

30. Varela JE, Wilson SE, Nguyen NT (2010) Laparoscopic surgery significantly reduces surgical-site infections compared with open surgery. Surgical endoscopy 24(2):270–276

31. Wichmann MW, Hüttl TP, Winter H, Spelsberg F, Angele MK, Heiss MM, Jauch KW (2005) Immunological effects of laparoscopic vs open colorectal surgery: a prospective clinical study. Archives of Surgery 140(7):692–697

32. Zitnick CL, Kang SB (2007) Stereo for image-based rendering using image over-segmentation. International Journal of Computer Vision 75(1):49–65

33. Zitova B, Flusser J (2003) Image registration methods: a survey. Image and vision computing 21(11):977–1000