Label-based Optimization of Dense Disparity Estimation for Robotic Single Incision Abdominal Surgery

V. Penza[1,2], S. Bacchini[2], A. Ciullo[2], E. De Momi[2], A. Forgione[3], L. S. Mattos[1]

[1]*Department of Advanced Robotics, Istituto Italiano di Tecnologia, Italy*

[2]*Department of Electronics Information and Bioengineering, Politecnico di Milano, Italy*

[3]*AIMS Academy, Milano, Italy*

veronica.penza@iit.it

INTRODUCTION

Minimally invasive surgical techniques have led to novel approaches such as Single Incision Laparoscopic Surgery (SILS), which allows the reduction of post-operative infections and patient recovery time, improving surgical outcomes. However, the new techniques pose also new challenges to surgeons: during SILS, visualization of the surgical field is limited by the endoscope field of view, and the access to the target area is limited by the fact that instruments have to be inserted through a single port.

In this context, intra-operative navigation and augmented reality based on pre-operative images have the potential to enhance SILS procedures by providing the information necessary to increase the intervention accuracy and safety [1]. Problems arise when structures of interest change their pose or deform with respect to pre-operative planning, as usually happens in soft tissue abdominal surgery. This requires online estimation of the 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 reconstruction, the more accurate the deformation identification can be.

This work presents an algorithm for 3D reconstruction of soft tissue, focusing on the refinement of the disparity map in order to obtain an accurate and dense point map. This algorithm is part of an assistive system for intra-operative guidance and safety supervision for robotic abdominal SILS [2].

MATERIALS AND METHODS

3D shape reconstruction using stereo-images is a process composed of two main steps: the disparity estimation and the stereo-triangulation. Using the left and right images (im_l, im_r) captured by a stereo camera, the disparity is estimated searching corresponding points between images. These are the projection of the same world point into two image planes. The relative distance between the corresponding pixels is described by the disparity map. Stereo-triangulation consists in exploiting the detected correspondence and the geometry of the cameras to extract the 3D measurement of the observed scene [3].

In this section, the steps of an improved algorithm for surface reconstruction are described (Fig. 1).

Disparity Estimation

Images are rectified in order to simplify the stereo-correspondence analysis. An intensity pixel-based algorithm is used for local similarity searching. A sparse modified census transform [4] is applied to im_l and im_r before matching. This method is robust against non-stationary exposure and illumination variations. It converts each pixel inside a moving window (nxn) into a bit vector, representing which neighbor pixels have an intensity above or below the central pixel and the mean of the pixels inside the window.

The similarity between the two census transformed images is done using the sum of Hamming Distance (HD) measure applied on a moving window. HD compares bit strings representing the pixels and identifies the number of positions at which the corresponding bits are different. These values are summed within a window (mxm) and a “winner-takes-all” strategy is used to find the minimum, and thus the disparity associated to the best similarity. The HD is computed only for a chessboard pattern of pixels inside the window in order to decrease the computational time.

Since using only pixel precision the reconstructed surface would consist of separated layers, a sub pixel refinement is applied using a parabola fitting. To invalidate wrong pixels on texture-less surfaces, the disparity is defined invalid if the two minimum values of HD are within a threshold. Additionally, a Left-Right Consistency (LRC) check is performed to invalidate half-occluded pixels, i.e. object views in one image and not in the other. A speckle removal filter is also applied in order to invalidate regions of large and small disparities that can be generated near the boundaries of the objects.

Dense Disparity Optimization

The pixels of the disparity map defined invalid lead to holes in the point cloud, compromising the usefulness of the reconstructed surface. Here, we exploit the segmentation of the reference image im_l to optimize the disparity map (Alg. 1). In literature, these techniques segment the images and then label each region with a disparity value, preserving the boundaries of the objects. However, finding gradually changing disparities can be a problem since all pixels in a region have the same disparity. Moreover, segmentation algorithms are computationally expensive.

***Algorithm 1** Label-based Disparity Optimization*

Require: D disparity map
L_i labeling of im_l segmented with SLIC Superpixel
M mask of invalid disparity values
τ threshold of number of valid points

```
associate L_i to D → DL_i
for each DL_i  i ∈ {1,..nSuperPixel}
    compute plane parameters p with LO-RANSAC of
    DL_i(i,j) with i = n_cols, j = n_rows
    compute median m of DL_i(i,j)
end for
for each i and j of D
    if M(i,j) = true
        if (nValidPoints ∈ DL_i) > τ
            replace D(i,j) with m (DL_i)
        else
            replace D(i,j) with a point on p(DL_i)
        end if
    end if
end for
return disparity map D_refined
```

To overcome these drawbacks, im_l is segmented using a Simple Linear Iterative Clustering (SLIC) super pixel algorithm [5]. Each super-pixel is an homogeneous area with similar or at least continuous depth. A label L is assigned to each super-pixel, being $im_l(L)$ a group of pixels with the same label, i.e. belonging to the same super-pixel. Only the invalid values of disparity map are corrected using the information of the pixels belonging to the same label. Depending on the valid values inside a label, we apply two different strategies:

- *Plane fitting*: if the found plane can be considered a reliable plane, the invalid disparity values are fitted to that plane. The plane parameters are estimated using the Locally Optimized RANSAC method.
- *Constant fitting*: otherwise, the invalid disparity values are replaced by the median of the valid ones.

After the stereo-triangulation step, a Moving Least Square algorithm is applied in order to smooth the point cloud, using PCL library.

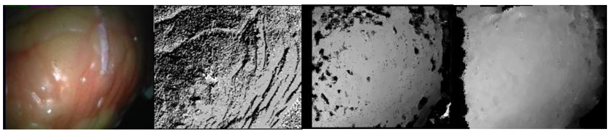

Fig. 1 Main steps of the presented algorithm. From left to right: RGB image, census transformed image, disparity map, dense refined disparity map.

RESULTS

The point clouds obtained are validated using 10 frames from a video dataset (*heart 1*) available online from the Hamlyn Centre [6], the closest available datasets to the application of this work. The error is defined as the median of Euclidean distances between the point cloud and the ground truth. We calculated this error for different window sizes of census transform and Hamming Distance to show the correlation between the accuracy and the computational time, as shown in Fig. 2. We considered also the percentage of valid points to evaluate the density of the point cloud (Fig. 2 – black line). Results show how the smoothing improves the accuracy of the reconstruction, giving a median error between 1.50mm and 2.27mm in the worst case.

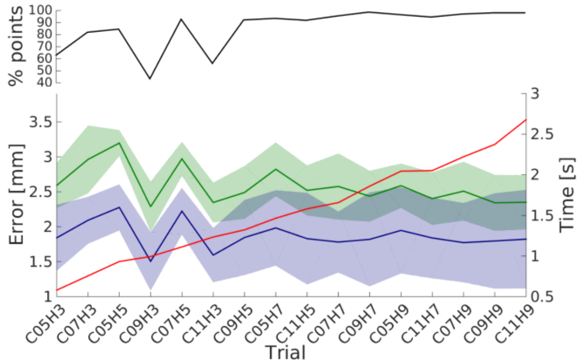

Fig. 2 Error before (green line) and after (blue line) the smoothing, and the pre-smoothing computational time (red line) with the corresponding interquartile range. C*n*H*m* represents the dimension of a squared window used in census transform (C) and in HD (H), ordered increasingly with time.

DISCUSSION

The presented work provides a dense surface reconstruction to be used in an enhanced vision system for robotic abdominal SILS. The strength of this method is the usage of SLIC Super Pixel algorithm for obtaining a high density valid disparity map (over the 90% of the total number of points), which can be exploited for augmented reality applications. Comparing our method with state-of-the-art CPU implementations [7] evaluated on *heart 1*, the percentage of valid points obtained with our method is 40% higher while providing a slightly lower accuracy (mean error 2.38 mm). Future research will focus on the development of a real-time implementation of the proposed algorithm, potentially based on a hybrid CPU-GPU processing framework.

REFERENCES

[1] Nicolau S, Soler L, Mutter D, Marescaux J. Augmented reality in laparoscopic surgical oncology. Surgical Oncology. 2011; 20(3):189-201
[2] Penza V, Ortiz J, De Momi E, Forgione A, Mattos L. Virtual assistive system for robotic single incision laparoscopic surgery. 4th Joint Workshop on Computer/Robot Assisted Surgery. 2014; 52-55
[3] Szeliski R. Computer vision: algorithms and applications. Springer Science & Business Media. 2010.
[4] Ma L, Li J, Ma J, Zhang H. A Modified Census Transform Based on the Neighborhood Information for Stereo Matching Algorithm. Image and Graphics (ICIG). 2013.
[5] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. Pattern Analysis and Machine Intelligence. 2012; 34(11): 2274-2282.
[6] Stoyanov D, Scarzanella M V, Pratt P, Yang G Z. Real-time stereo reconstruction in robotically assisted minimally invasive surgery. MICCAI. 2010; 275-282
[7] Röhl S, et al. Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration. Medical Physics. 2012; 39(3): 1632-1645.