Registration of 2D monocular endoscopy to 3D CBCT for video-assisted thoracoscopic surgery

Baptiste Noblet, Matthieu Chabanas, Simon Rouzé, Sandrine Voros

Registration of 2D monocular endoscopy to 3D CBCT for Video-Assisted Thoracoscopic Surgery

Baptiste Noblet[a], Matthieu Chabanas[a], Simon Rouzé[b], and Sandrine Voros[a, c]

[a]Université Grenoble Alpes, CNRS, Grenoble INP, TIMC; F-38000 Grenoble, France
[b]CHU Rennes, Department of Cardio-Thoracic and Vascular Surgery; F-35000 Rennes, France
[c]INSERM; F-38000 Grenoble, France

ABSTRACT

Cone-Beam CT (CBCT) is a valuable imaging modality for the intraoperative localization of pulmonary nodules during Video-Assisted Thoracoscopic Surgery (VATS). However, inferring the nodule position from the CBCT to the operative field remains challenging and could greatly benefit from computer-aided guiding. As a first step towards an Augmented Endoscopy guiding system, we propose to register 2D monocular endoscopic views into the 3D CBCT space. Ribs and wound protectors are segmented in both imaging modalities, then registered using an image-to-cloud Iterative Closest Point variant. The method is evaluated qualitatively on clinical VATS video sequences from 3 patients. The promising results validate this first step towards a seamless monocular VATS navigation.

Keywords: 2D-3D registration, Iterative Closest Point, CBCT, Monocular Endoscopy, Augmented Endoscopy, Video-Assisted Thoracoscopic Surgery

1. INTRODUCTION

Video-assisted thoracoscopic surgery (VATS) is a minimally invasive surgical technique for the diagnosis and treatment of early-stage lung cancer. Nodule resection is performed through incisions in the thorax under the control of an endoscopic view, in which the nodule is not visible (Figure 1: Right). A major challenge is that the intraoperative collapsed lung is severely deformed with respect to the preoperative CT-scan, due to the patient change of pose from supine to lateral decubitus and due to the induced pneumothorax (lung deflation). These deformations yield a high uncertainty in the actual nodule localization. This can be seen in Figure 1 as the preoperative CT on the left and the intraoperative image in the middle show the same nodule. Standard nodule localization strategies are based on the placement of markers such as hook-wires, micro-coils or dyes[1,2]. However these strategies require an additional pre- or intraoperative procedure with possible complications which significantly increase the length and complexity of the surgery.

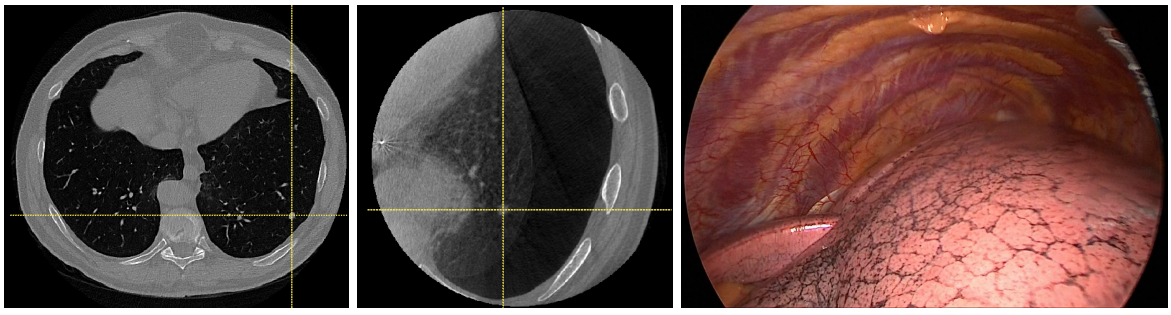

Figure 1: Left: Preoperative CT showing the lung nodule. Middle: Intraoperative CBCT showing the same nodule after lung deformation. Right: Endoscopic image where the nodule is not visible.

Intraoperative imaging has become an alternative to preoperative marking using ultrasound[3] or Cone Beam CT (CBCT)[4,5]. When the nodule is difficult to localize, several registration methods have been proposed to map

the preoperative nodule location on US or CBCT intraoperative images, and have been evaluated on animals[6] or retrospective clinical cases[7,8]. A remaining challenge once the nodule is localized in the intraoperative images is to guide the surgeon in the operative field. To our knowledge, only Rouzé *et al.*[4] have described their use orthogonal fluoroscopic images to position a surgical instrument close to the nodule. However this procedure remains cumbersome as information is displayed on the CBCT screen in the hybrid room and not on the endoscope screen. Furthermore, there is a high exposure to radiation for the surgical team since the instruments must be held manually during fluoroscopy.

We have started to develop an Augmented Endoscopy guiding system where the nodule localization estimated in the CBCT is directly overlaid on the endoscopic view. This system would remove the need for fluoroscopy, simplify the procedure and greatly enhance the comfort and safety of the surgical team. While Augmented Endoscopy has been proposed for other applications[9–11], it has to our knowledge never been studied for VATS, except for 2D tracking of manually placed anatomical points on the surface of the lung[12]. Although our setup requires a CBCT, typically in an hybrid Operating Theater (OR), one of our objectives is to avoid the use of non-standard devices such as a stereo endoscope or a tracking system.

A complete setup for real time endoscopic augmentation will require registering the CBCT and endoscopic spaces, tracking the camera motion inside the rib cage and compensating the lung movements. In this paper, we focus on the first step, namely the registration of 2D endoscopic views to the 3D CBCT.

Several approaches are possible to solve this task[9]. The most likely to work with our data are the following, succinctly presented with their limitations:

- Using a tracking device, such as a NDI Polaris.

  By tracking the endoscope and the C-arm, the transformation between the endoscopic image and the nodule localization in the 3D CBCT can be computed and the nodule position can be projected onto the video.

  However, those types of systems require installation, calibration and sterilization of some parts, which is a significant modification of the clinical protocol.

- Leaving the endoscope inside the CBCT during acquisition[13].

  By reconstructing the tip of the endoscope in the CBCT, it is possible to estimate the transform between the camera position and the nodule position.

  However, this method does not seem clinically feasible as the endoscope needs to be steadied with a mechanical arm, which could easily collide with the C-arm during acquisition.

- Registering a video-based 3D reconstruction of the thoracic cavity with the CBCT.

  Visual odometry methods can be used to reconstruct surfaces in scenes and has been employed for Minimally Invasive Surgery sequences.

  However, most of those methods require a rigid scene[14], a stereo endoscope[15] or do not work as some of their hypothesis(planar scene,[16] slow deformations[17] , ...) are not valid in our context.

- 2D-3D registration using features visible in both modalities.

  By pairing features between modalities, a camera pose can be estimated. Those features can be natural, surgical ancillaries or additional equipment.

  Adding additional equipment such as surgical clips is problematic and few features are visible in CBCT and endoscopic video in VATS.

We chose a feature-based method, using ribs and wound protectors which are natural features and standard surgical ancillaries, respectively. Since there features are always present in the surgical scene, we limit the surgical workflow disruptions, making it easier to adopt and allowing evaluation on retrospective clinical data. Both features are visible in each modalities (Figure 2), however both have difficulties that need to be addressed, especially their extraction as they can be partially hidden or missing.

We propose to first segment the ribs and wound protectors from the CBCT and the endoscopic images, then register these features using a modified 2D-3D iterative approach. The proposed method is evaluated on three retrospective clinical cases. Promising qualitative results on clinical cases are presented.

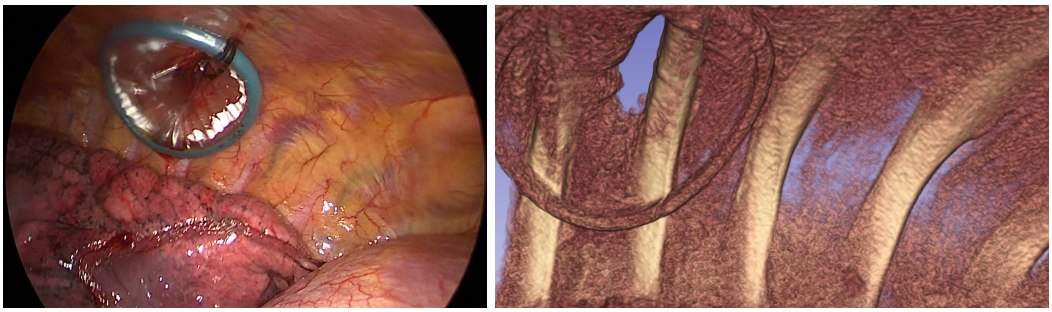

Figure 2: Left: Example of endoscopic video frame. Right: Volume rendering of the CBCT of approximately the same region. Wound protectors and ribs can be seen in both modalities.

2. MATERIAL AND METHODS

2.1 Material

Endoscopic images were acquired with a Stryker® RGB HD monocular camera mounted on a 10mm Stryker® 30° endoscope.

Camera calibration was done using standard computer vision methods with OpenCV. The calibration rig is an A4 size, 4x4 markers, 10x8 Charuco[18] board. Camera parameters of the endoscope, especially the zoom factor, were kept constant to avoid the need for recalibration. The calibration procedure took around 30 seconds for the surgeon, which consisted of keeping the endoscope steady while the calibration rig was slowly moved in front of it. This process is short enough to be compatible with clinical use. The calibration computation was done using OpenCV and a random sample of 50 frames outputs a reprojection error of less than a pixel.

The three incisions are protected with two XXS and one XS Alexis® "O Wound Protector/Retractor" from Applied Medical®. The distal ring of these covers can be seen in both the CBCT and the endoscope.

Following Rouzé *et al.*[4] clinical protocol, we acquired intraoperative images with a GE® Discovery™ CBCT in a hybrid OR. Since the device Field of View (FOV) is smaller than the thoracic cavity, the image volume was centered on the operated lobe as seen Figure 1 (middle).

2.2 Overview of the method

Our method is a 2D-3D registration using on wound protectors and ribs as features. It is based on an ICP[19] variant for 2D-3D registration,[20] using segmentation of the features and modified specifically for VATS.

The segmentation of both features are presented in section 2.3 and the pairing between modalities in section 2.4.5.

Section 2.4.1 presents modifications added to [20] such as instance matching, namely each rib and wound protector are individually paired and accounted for in the registration cost function and sections 2.4.2 and 2.4.3 presents features specific to multi-port surgery and VATS such as a constrained optimization to ensure that the estimated endoscope goes through the wound protector.

The endoscope typically goes through one specific port during VATS. However, the surgeon may want to change the point of view. Section 2.4.4 describes how the entry port is automatically detected.

2.3 Feature segmentation

Monocular RGB video and 3D CBCT images are too different to consider purely image-based registration approaches. To circumvent this difficulty, the registration is conducted on the segmentation of the features in both modalities.

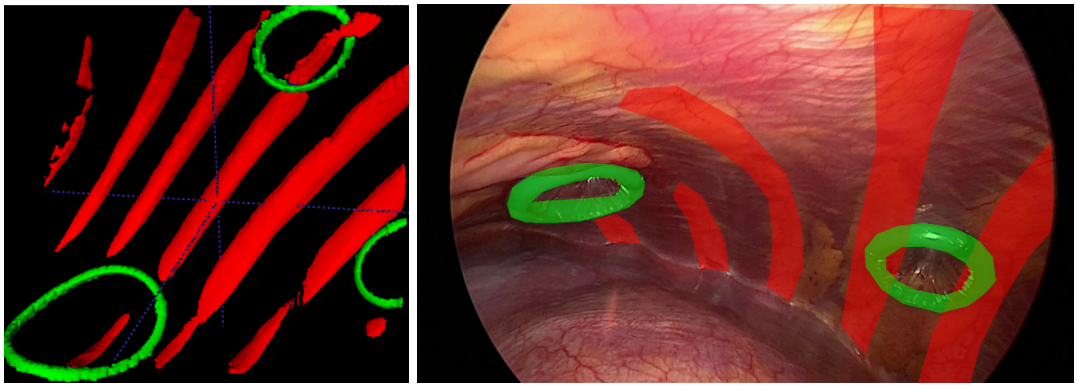

Figure 3: Left: Example of 3D segmentation. Ribs are only partially visible in the CBCT because of the limited CBCT FOV. Right: Example of 2D segmentation. Ribs are difficult to outline and can be hidden by adipose tissues.

2.3.1 Ribs

Ribs are segmented automatically in the CBCT with classical image processing methods and morphological operations. Operations are applied on a smoothed out version of the CBCT to remove most of the noise, using a gaussian kernel of 1mm. The bones are segmented with a high threshold (HU:200) and closing morphological operations. Instead of using the whole rib as the feature, we only want to use the part visible by the camera. By intersecting the rib mask with a dilated air mask, which is obtained with a low threshold (HU:-650), we select only the part of the rib that is close to the parietal pleurae, which corresponds to the visible portion of the rib in the endoscopic video. This outputs the thin portion of the ribs (Figure 3 left).

Ribs are difficult to segment in the endoscopic video, even manually, as can be seen on the left image in Figure 2. The ribs visibility is highly dependent on the patient's morphology and rib cage section. Adipose tissue (yellow tissue) can appear on the parietal pleurae and the outline is not visible as ribs gradually disappear under intercostal muscle. Because of those difficulties, the segmentation is currently done manually using CVAT*. A simple polygon is used to describe approximately the outer shape of each ribs, and a surface mask is generated from it, with uncertainty values increasing closer to the outline. The uncertainty map is obtained by smoothing the mask and setting to zero values outside of the mask. An example of manual segmentation is shown on the right in Figure 3.

In both modalities, ribs are separated in connected components. A rib may be separated into two or more components, for example cut by the CBCT FOV (Figure 3 Left) or partially hidden in the endoscopic video by a wound protector (Figure 3 Right).

2.3.2 Wound protectors

The wound protectors used during the surgery are "O Wound Protector/Retractor" (Alexis®), which are two plastic torus linked by a thin plastic sheet. Wound protectors are easily and quickly segmented manually in the endoscopic video as they have a very distinct blue color. The segmentation is done using CVAT by tracing two low resolution polygon to describe the inner and outer shapes of the wound protector blue ring.

However, their segmentation in the CBCT is not simple as their Hounsfield values are close to flesh values and they may be partially or completely outside of the CBCT field of view. The Hounsfield value of each element is experimentally indistinguishable from flesh. Segmentation is currently done manually using ITKsnap† adaptive brush with additional smoothing. Wound protectors can be partially visible in the CBCT. This can be detected and corrected by extending them artificially with their real dimensions. Visibility detection is done by checking the amount of surface touching the boundary of the CBCT FOV. Visibility extension is done by registering a

*https://www.cvat.ai/

†http://www.itksnap.org

synthetic volume representation of the full wound protector using the Elastix[‡] framework. In cases where the wound protector is outside of the CBCT FOV, a segmentation can still be acquired when done on the extended reconstruction of the CBCT. This reconstruction is obtained using RTK[21] and the projection images obtained by the C-arm. Otherwise, the port is discarded and not used during the registration process.

Wound protectors are separated and treated as unique entities in both modalities.

2.4 Registration

The registration method is a variant of the ICP algorithm[19] for 2D/3D registration, similarly to [20]. However several modifications were added to [20] in order to work within our context. Modifications that can be applied to another context are presented in sections 2.4.1 and 2.4.6 and VATS specific modifications in sections 2.4.2,2.4.3, 2.4.4 and 2.4.5.

2.4.1 2D/3D Iterative Closest Point

The general principle behind the method is the same as the classical ICP:

1. Find the correspondence minimizing the distance between the sets of points.
2. Find the optimal transform to minimize that distance with this correspondence.
3. Iterate until convergence or for a number of iterations.

We register the whole segmentation instead of the contours only, which removes the need for the feature detection of [20]. This is because our 2D segmentation of the rib has high uncertainty on the edges. Each segmentation is sampled and converted to point cloud. The size of the cloud of each instance is proportional to the number of elements in the segmentation with the biggest rib having 1000 elements in both modalities.

During the matching phase (step 1), each point cloud is treated as independent, but during the optimization phase (step 2), only one transformation is computed, as if all clouds were concatenated into one. All errors are summed to determine the total error. Each feature's error is adjusted by a function to modulate its influence during the optimization. The distance used is the euclidean distance between the 3D point and the projection line of the associated 2D projection, weighted by the uncertainty map.

Three steps are added to remove aberrant matches on a per component basis, with a threshold set to 3 times the median distance for each component:

- 3D points outside of the camera FOV.
- Removal of high 2D distance after projection to remove points that are close to the camera center.
- Removal of high 3D point-to-line distance matches to allow some inaccuracies in the segmentation.

2.4.2 Feature weighting

The VATS context can be further used to constrain and help the registration process.

The similarity between adjacent ribs and the alignment uncertainty along their lengths can easily result in a computed registration shifted by a least an intercostal distance (around 25mm) or a rotational error around the spine axis. The small size, flexible and circular shape of the wound protectors make them unreliable for accurate rotation estimations. By using the wound protectors as anchors to prevent a shift between ribs and along their lengths, and by using the ribs as the main registration feature, we expect that using both features outputs better and more robust results. This is done by modulating the distance of the wound protectors. By overestimating the distances when they are not properly aligned, we give them more influence at the beginning of the registration process. When we are sure that the ribs are not misaligned, their influence is diminished. This is done by squaring the wound protectors errors.

[‡] https://elastix.lumc.nl/

2.4.3 Constrained optimization

The endoscopic camera goes through a port. This can be taken into account by parametrizing the optimizer vector to only compute physically possible positions. The endoscope is still allowed its 6 Degrees-of-Freedom but the port-based coordinates allow fine control and bounding of the possible values, preventing it from leaving the port or going inside surrounding tissue. A Principal Component Analysis is computed on the entry wound protector segmentation and used to fit an ellipse. This ellipse provides a simple model describing the wound protector and provides a basis for the coordinate system.

2.4.4 Entry port

The endoscope is usually placed through the lowest of the three ports, on the medo-axillary line. However the surgeon may want to view the scene from a different angle by inserting it through another port. This means that no assumption should be made about the port which contains the endoscope. In order to determine which port is the correct one, we register all possible configurations and select the best one. The criteria is the lowest registration error after a few iterations. With three ports, there are six possible configurations that can be tested in parallel:

- If two wound protectors are visible by the camera: all port can be the point of entry and for each, two permutations of wound protectors exist.
- If one wound protector is visible by the camera: each of the three wound protector can be the one seen from two different entry points.

The endoscope is initially placed in the middle of the entry port, looking at the other wound protector. Because of the constraints of the endoscope staying through the port, the optimization process tries to remove the endoscope from the patient without succeeding on wrong configurations, resulting in aberrant registration with a high error. This allows incorrect configurations to be detected and discarded. After the determination of the best configuration, it is fully optimized to find the accurate registration.

2.4.5 Instance matching

Ribs and wound protectors can be treated as a single element. However we use the fact that we can segment each rib and wound protector independently to help the registration. Each instance of a feature must be matched to its corresponding element or discarded if it is not present in the other modality.

By doing so, we prevent some aberrant situations such as two sides of a single rib matching with two different ribs in the other modality. However this pairwise matching is not straightforward as the ribs in the CBCT and in the video are only partially visible and not necessarily the same section. We sort the ribs in both modalities and give them a value. Each rib is separated by an intercostal distance which allows us to number them, determine if one is missing and combine components of the same rib. This outputs two unpaired sets of ribs. The visible wound protectors, which are already matched based on the configuration, are then used to find the matching between the two sets of ribs. An example of matched ribs between modalities is shown in Figure 4.

In case sorting or matching the ribs failed, a partial registration is computed, with every rib considered as a single element. Then the ribs are paired based on the current estimation of the camera pose and the closest rib in the other modality.

2.4.6 Implementation details

Because of the distance choice and the fact that all projection lines converge to a unique point (the camera position), the matching between 3D points and their closest projection lines can be efficiently implemented. A 2D nearest neighbor data structure, such as a KDTree, is created to represent the points segmented in the image. To find the correspondence, the 3D points are projected onto the current camera position estimation. Each point is matched to the closest 2D image point and then the 3D distance is computed based on this matching. This is an efficient matching technique as the tree only depends on the input 2D image, so it is computed once and only queried during iterations.

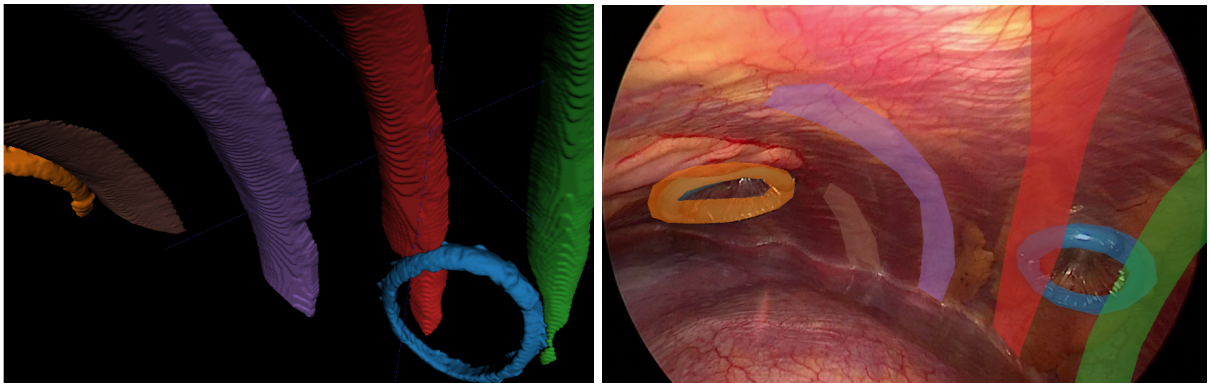

Figure 4: Example of matched pairs shown by color. Ports (orange and blue) are paired based on the configuration, allowing ribs to be sorted and matched.

As we want to constrain the endoscope (Section 2.4.3) inside the port during the optimization, we cannot use the Levenberg-Marquardt algorithm as in [20]. Instead, we use the Trust Region Reflective algorithm, implemented in scipy `optimize.least_square`. The number of function calls is limited to a small number (32) to lower the time between new closest correspondence searches. This greatly decreases the computation time by not fully optimizing correspondences that are no longer optimal.

3. EXPERIMENTS

Three patients (two right lungs and one left lung) were operated with the procedure described by Rouzé *et al.*[4] The resections were performed in a hybrid operating room at the Rennes University Hospital, France, under the approval of the local ethics committee (2016-A01353-48 35RC16 9838) and the informed consent of all patients.

Panoramic views of the inside of the rib cage were acquired from several ports when possible, which amounts to 5 sequences. 7 frames were segmented manually and registered.

Of the 9 wound protectors spread among the 3 patients, 6 wound protectors were fully inside the CBCT, 1 was partially present and reconstructed and 1 was recovered with the extended CBCT and 1 was too low to be visible.

4. RESULTS

Figure 5 shows the initialization and result of the registration for a specific frame. The entry port was correctly identified and the registration process converged. The remaining errors can be attributed to the limited segmentation of each modality, leading to a poor overlap. For example the rib on the left(brown) does not have any overlap in both modalities due to limited CBCT FOV and adipose tissue.

A volume rendering of the CBCT from the computed transform (Figure 6 Left) shows that the computed transform is qualitatively satisfying.

The endoscope can be reconstructed inside the CBCT (Figure 6 Right) and it shows that the endoscope stays through the port, despite being able to go further as it is limited to stay inside the visible wound protector ring.

In total, 5 frames from different patients were successfully registered while 2 failed to identify the correct configuration (one with a single wound protector visible and one with only a partial view of the wound protector in the video).

For the moment, the computation time is of a few seconds per frame, using python.

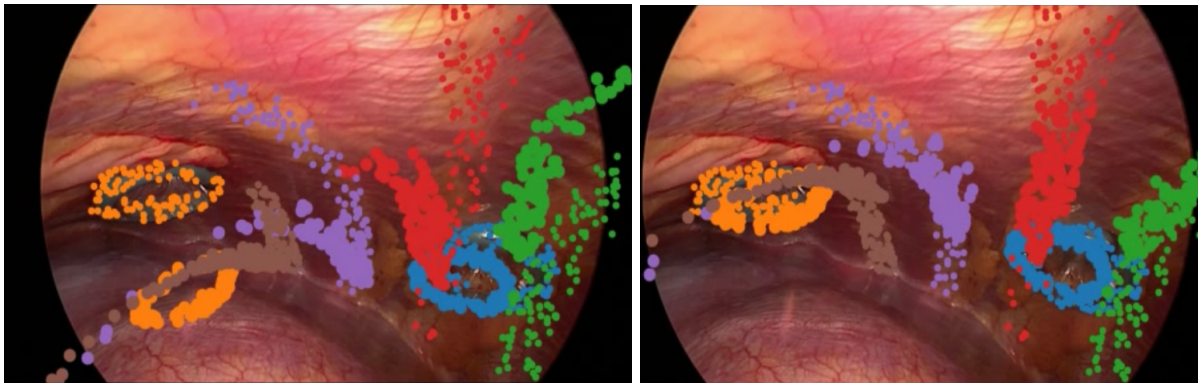

Figure 5: Example of registration result. Both 2D and 3D segmentation are shown. Left: Initialization of the right configuration; the two points clouds do not align. Right: Output of the registration. Most features are properly registered, apart from the brown rib which still has a residual error because of the lack of overlap between modalities.

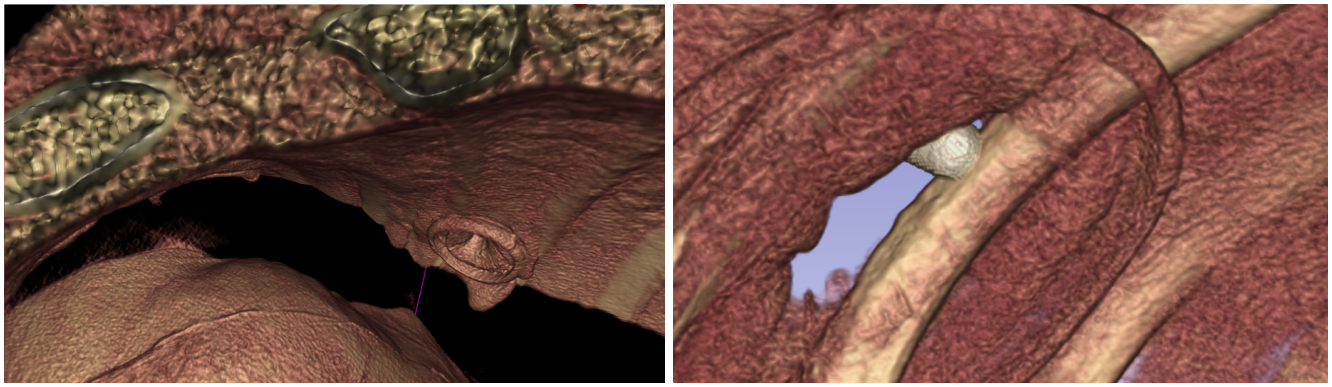

Figure 6: Left: Volume rendering of the CBCT from the computed camera pose which aligns with Figure 5. Right: Reconstructed endoscope. The endoscope stays inside the port even when allowed as far as the plastic ring.

5. DISCUSSION AND CONCLUSION

The clinical protocol to calibrate the camera and acquire the images is easy and fast for the surgeon, and thus compatible with a clinical routine. No input is needed to specify the port through which the endoscope goes. The instructions given to the surgeon are to pan slowly to avoid motion blur, make sure that a decent amount of rib cage is visible and at least one wound protector. Those instructions are easy to remember and execute, not hindering the surgeon's concentration. Limiting the surgeon's ability to change the zoom during the augmentation does not seem to be an issue. Some parts of the method can be more robust by making sure that the wound protectors and ribs are mostly in the CBCT field of view but it is not a critical part. The method is not yet viable for real time augmentation but is sufficient for near static augmentation, when the endoscope is kept steady for a few seconds.

As we worked on retrospective data, we have no quantitative method for evaluation or validation yet. However, qualitative results show that the method is promising for this application. Possible methods for quantitative evaluations are being investigated for future patients, such as the use of surgical clips for Perspective-n-Points registration methods, stereo endoscopy or tracker-based methods.

Future works on this method will mostly focus on the automation of the segmentation, eventually with the segmentation outside the CBCT FOV, either from the extended CBCT or the registered preoperative CT. A possible approach to solve the segmentation of the wound protectors in the CBCT is to use morphological

operations. For the segmentation of the ribs in the endoscopic video, we found experimentally that adipose tissue tend to follow the rib line, making them feasible to segment.

This method is the first component of a future automatic augmented reality navigation system for nodule resection. These results show us that our monocular and tracker-less approach seems feasible. We will now address the remaining steps towards a complete Augmented Reality guiding system for VATS.

ACKNOWLEDGMENTS

This work has been partly supported by the French National Research Agency through the frameworks ANR-11-LABX-0004 and ANR-20-CE19-0015.

REFERENCES

[1] Keating, J. and Singhal, S., "Novel methods of intraoperative localization and margin assessment of pulmonary nodules," *Semin Thorac Cardiovasc Surg* **28**(1), 127–136 (2016).

[2] Velasquez, R., Martin, A., Abu Hishmeh, M., DeLorenzo, L., Dhillon, S. S., and Harris, K., "Placement of markers to assist minimally invasive resection of peripheral lung lesions," *Annals of Translational Medicine* **7**, 360–360 (aug 2019).

[3] Wada, H., Anayama, T., Hirohashi, K., Nakajima, T., Kato, T., Waddell, T. K., Keshavjee, S., Yoshino, I., and Yasufuku, K., "Thoracoscopic ultrasonography for localization of subcentimetre lung nodules," *European Journal of Cardiothoracic Surgery* **49**, 690–697 (Feb. 2016).

[4] Rouzé, S., de Latour, B., Flécher, E., Guihaire, J., Castro, M., Corre, R., Haigron, P., and Verhoye, J.-P., "Small pulmonary nodule localization with cone beam computed tomography during video-assisted thoracic surgery: a feasibility study," *Interactive CardioVascular and Thoracic Surgery* **22**(6), 705–711 (2016).

[5] Saito, Y., Watanabe, T., Kanamoto, Y., Asami, M., Yokote, F., Dejima, H., Morooka, H., Ibi, T., Yamauchi, Y., Takahashi, N., Ikeya, T., Sakao, Y., and Kawamura, M., "A pilot study of intraoperative localization of peripheral small pulmonary tumors by cone-beam computed tomography: sandwich marking technique," *Journal of Thoracic Disease* **0**(0) (2022).

[6] Uneri, A., Nithiananthan, S., Schafer, S., Otake, Y., Stayman, J. W., Kleinszig, G., Sussman, M. S., Prince, J. L., and Siewerdsen, J. H., "Deformable registration of the inflated and deflated lung in cone-beam CT-guided thoracic surgery: Initial investigation of a combined model- and image-driven approach," *Medical Physics* **40**(1), 017501 (2013).

[7] Lesage, A.-C., Rajaram, R., L Tam, A., Rigaud, B., K Brock, K., C Rice, D., and Cazoulat, G., "Preliminary evaluation of biomechanical modeling of lung deflation during minimally invasive surgery using pneumothorax computed tomography scans," *Physics in Medicine & Biology* **65**, 225010 (Nov. 2020).

[8] Alvarez, P., Rouzé, S., Miga, M. I., Payan, Y., Dillenseger, J.-L., and Chabanas, M., "A hybrid, image-based and biomechanics-based registration approach to markerless intraoperative nodule localization during video-assisted thoracoscopic surgery," *Medical Image Analysis* **69**, 101983 (2021).

[9] Bernhardt, S., Nicolau, S. A., Soler, L., and Doignon, C., "The status of augmented reality in laparoscopic surgery as of 2016," *Medical Image Analysis* **37**, 66–90 (2017).

[10] Simpfendörfer, T., Baumhauer, M., Müller, M., Gutt, C. N., Meinzer, H.-P., Rassweiler, J. J., Guven, S., and Teber, D., "Augmented reality visualization during laparoscopic radical prostatectomy," *Journal of endourology* **25**(12), 1841–1845 (2011).

[11] Khaddad, A., Christophe, J., Margue, G., Michiels, C., Ricard, S., Chandelon, K., Bladou, F., Bourdel, N., and Bartoli, A., "A survey of augmented reality methods to guide minimally invasive partial nephrectomy," *World Journal of Urology* (2022).

[12] Thienphrapa, P., Bydlon, T., Chen, A., Vagdargi, P., Varble, N., Stanton, D., and Popovic, A., "Interactive Endoscopy: A Next-Generation, Streamlined User Interface for Lung Surgery Navigation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11768 LNCS**, 83–91 (2019).

[13] Bernhardt, S., Nicolau, S. A., Agnus, V., Soler, L., Doignon, C., and Marescaux, J., "Automatic localization of endoscope in intraoperative CT image: A simple approach to augmented reality guidance in laparoscopic surgery," *Medical Image Analysis* **30**, 130–143 (2016).

[14] Wu, H., Xu, R., Xu, K., Zhao, J., Zhang, Y., Wang, A., and Iwahori, Y., "3D Texture Reconstruction of Abdominal Cavity Based on Monocular Vision SLAM for Minimally Invasive Surgery," *Symmetry* **14**, 185 (jan 2022).

[15] Zhou, H. and Jayender, J., "Emdq-slam: Real-time high-resolution reconstruction of soft tissue surface from stereo laparoscopy videos," in [*Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*], 331–340, Springer (2021).

[16] Gomez-Rodriguez, J. J., Lamarca, J., Morlana, J., Tardos, J. D., and Montiel, J. M. M., "SD-DefSLAM: Semi-Direct Monocular SLAM for Deformable and Intracorporeal Scenes," 5170–5177 (2021).

[17] Rodriguez, J. J. G., Montiel, J. M. M., and Tardos, J. D., "Tracking monocular camera pose and deformation for SLAM inside the human body," (2022).

[18] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., and Marín-Jiménez, M. J., "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition* **47**(6), 2280–2292 (2014).

[19] Besl, P. J. and McKay, N. D., "Method for registration of 3-d shapes," in [*Sensor fusion IV: control paradigms and data structures*], **1611**, 586–606, Spie (1992).

[20] Pujol-Miro, A., Ruiz-Hidalgo, J., and Casas, J. R., "Registration of images to unorganized 3D point clouds using contour cues," *25th European Signal Processing Conference, EUSIPCO 2017* **2017-Janua**, 81–85 (2017).

[21] Rit, S., Oliva, M. V., Brousmiche, S., Labarbe, R., Sarrut, D., and Sharp, G. C., "The reconstruction toolkit (rtk), an open-source cone-beam ct reconstruction toolkit based on the insight toolkit (itk)," in [*Journal of Physics: Conference Series*], **489**(1), 012079, IOP Publishing (2014).