



Automatic localization of endoscope in intraoperative CT image: A simple approach to augmented reality guidance in laparoscopic surgery[☆]



Sylvain Bernhardt^{a,b,*}, Stéphane A. Nicolau^{c,**}, Vincent Agnus^c, Luc Soler^{a,c},
Christophe Doignon^b, Jacques Marescaux^{a,c}

^aIHU, Institut Hospitalo-Universitaire, Strasbourg, France

^bICube (UMR 7357 CNRS), Université de Strasbourg, France

^cVirtual Surg, IRCAD, Strasbourg, France

ARTICLE INFO

Article history:

Received 19 August 2014

Revised 17 April 2015

Accepted 4 January 2016

Available online 13 February 2016

Keywords:

Augmented reality

C-arm

Registration

Intraoperative CT

ABSTRACT

The use of augmented reality in minimally invasive surgery has been the subject of much research for more than a decade. The endoscopic view of the surgical scene is typically augmented with a 3D model extracted from a preoperative acquisition. However, the organs of interest often present major changes in shape and location because of the pneumoperitoneum and patient displacement. There have been numerous attempts to compensate for this distortion between the pre- and intraoperative states. Some have attempted to recover the visible surface of the organ through image analysis and register it to the preoperative data, but this has proven insufficiently robust and may be problematic with large organs. A second approach is to introduce an intraoperative 3D imaging system as a transition. Hybrid operating rooms are becoming more and more popular, so this seems to be a viable solution, but current techniques require yet another external and constraining piece of apparatus such as an optical tracking system to determine the relationship between the intraoperative images and the endoscopic view. In this article, we propose a new approach to automatically register the reconstruction from an intraoperative CT acquisition with the static endoscopic view, by locating the endoscope tip in the volume data.

We first describe our method to localize the endoscope orientation in the intraoperative image using standard image processing algorithms. Secondly, we highlight that the axis of the endoscope needs a specific calibration process to ensure proper registration accuracy. In the last section, we present quantitative and qualitative results proving the feasibility and the clinical potential of our approach.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

By the end of the last century, surgeons had started to change the way they operate from simply cutting open their patients to performing the intervention through small skin incisions or even natural orifices. This minimally invasive surgery has many advantages compared to traditional open surgery, such as smaller scars, shorter recovery time and fewer complications. However, the surgeon's loss of direct vision of the organs compelled them to use an

endoscopic camera for guiding their instruments and actions. As the intervention is performed via a digital video display, the possibility emerges to augment the rendered scene with additional data relevant to the operation, thus improving the surgeon's experience. This idea of creating Augmented Reality (AR) systems for guidance in minimally invasive surgery has been the focus of a lot of research in recent years (Sielhorst et al., 2008).

Usually, AR in laparoscopic surgery is achieved by registering a preoperative high definition 3D model (e.g. segmentation from CT or MRI images) with the intraoperative imaging modality (commonly the video display from the endoscopic camera) (Baumhauer et al., 2008; Nicolau et al., 2011). In abdominal surgery, a pneumoperitoneum is realized by insufflating gas, creating enough room in the abdominal cavity for the insertion and motion of the surgical instruments below the skin surface. However, the pressure applied by the pneumoperitoneum makes the organs change

[☆] This paper was recommended for publication by James Duncan.

* Corresponding author.

** Corresponding author.

E-mail addresses: sylvain.bernhardt@ihu-strasbourg.eu (S. Bernhardt), stephane.nicolau@ircad.fr (S.A. Nicolau).

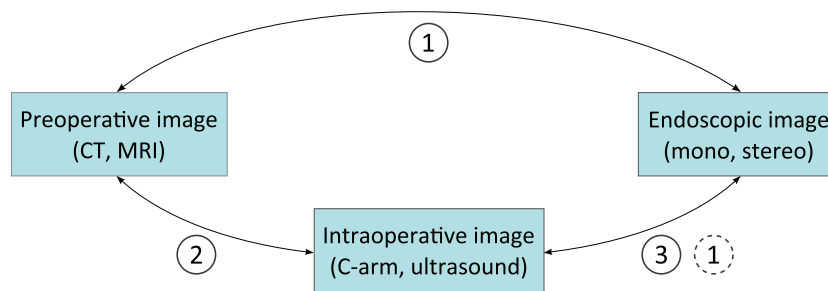


Fig. 1. Diagram illustrating how the different approaches are positioned with respect to augmented reality purposes. ① represents the methods based on Structure from Motion (SfM) or stereovision. ② stands for the methods aiming at determining the anatomical deformation occurring between the preoperative and the intraoperative states, notably due to the pneumoperitoneum. ③ symbolizes all the techniques to locate the endoscopic camera with respect to the intraoperative scene. The latter is also inherently achieved by ①, hence the dashed circle.

their position and shape. In addition to the patient's displacement between the preoperative acquisition and the intraoperative state, this strongly hinders an accurate registration (Sánchez-Margallo et al., 2011). Although a rigid registration method may provide some efficient assistance to surgeons (Marescaux et al., 2004; Nicolau et al., 2013; Mårvik et al., 2004), it is clear that an acceptable solution should routinely compensate for this deformation to ensure sufficient accuracy. So far, a number of different approaches have been developed towards this end and are summarized in the Fig. 1.

The first approach (① in Fig. 1) is one of the methods that aim at extracting the visible surface of the organ with optical techniques (Maier-Hein et al., 2013) such as Structure from Motion (SfM) (Pizarro and Bartoli, 2012; Mahmoud et al., 2012), stereovision (Lo et al., 2008; Hu et al., 2009; Mountney and Yang, 2010; Haouchine et al., 2013) or Shape from Shading (Malti and Bartoli, 2014). Then the extracted mesh is registered in a non-rigid way with the corresponding surface segmented from preoperative images. The position of the camera with respect to the surface is inherently calculated using techniques such as SfM and stereovision. Therefore, no external tracking system is required. Such a process can allegedly be executed in near real-time (Figl et al., 2010; Su et al., 2009), but the registration is reliable only on the visible surface. This could be a problem for a large organ such as the liver, which can present a significant deformation discrepancy. Moreover, the method requires some anatomical landmarks in order to remove any uncertainty in the registration process and serve as a referential frame for the camera localization. This makes all the estimations interdependent and therefore less robust. Finally, the lack of intraoperative ground truth for organ surfaces prevents a proper accuracy validation.

The second approach is to split the whole registration into two distinctive parts by introducing intraoperative imaging as an intermediary stage (see Fig. 1). First, an attempt is made at estimating the complete abdomen deformation using biomechanical models (Kitasaka et al., 2004; Bano et al., 2012; Oktay et al., 2013) or intraoperative low-dose CT acquisitions (Feuerstein et al., 2008; Shekhar et al., 2010; Bano et al., 2013). Then, the preoperative images are updated according to the estimated deformation and registered to the intraoperative scans (② in Fig. 1). The second part is to use an external tracking system to perform the extra registration step of aligning the intraoperative images with the endoscopic video (③ in Fig. 1).

There exist several ways of determining an endoscopic camera position and motion. In Mårvik et al. (2004), a rigid registration was performed using radio-opaque markers stuck onto the patient's skin, extracted in a preoperative CT image and then picked on the skin using a pointer visually tracked by an external system using tags. However, the method was too constrictive and inaccurate due to the approximate pointing system and the defor-

mation from the pneumoperitoneum. Ever since, optical tracking has become the predominant solution for endoscope and instrument tracking, as illustrated in Feuerstein et al. (2008). While the accuracy is satisfactory, the technique is costly, cumbersome and requires many calibration steps that can slow down the surgical workflow (Shahidi et al., 2002; Garcia et al., 2004; Nicolau et al., 2005). Moreover, the infrared reflectors attached to the tracked devices have to be in sight of the cameras at all times, constraining the movements of the surgeon.

So, a wide variety of methods have been developed over the last decade around the two approaches illustrated in Fig. 1. However, none seems to fully bring a solution of augmented reality in minimally invasive surgery. Indeed, the first approach lacks robustness and is too constrictive as it relies solely on the visible surface. The second approach thus has greater appeal, especially since, in recent years, the 3D rotational C-arm is becoming increasingly popular and is sometimes even mandatory for certain interventions (cardiovascular surgery and neurosurgery). Its relatively small bulk enables the use of high quality intraoperative imaging (Feuerstein et al., 2008) and it is steadily making its way into an increasing number of operating rooms. Nevertheless, the determination of the 3D endoscope pose with respect to the intraoperative image (③ in Fig. 1) requires an extra piece of apparatus in the surgical site and additional calibration steps that lengthen the intervention.

In light of the above, we propose a new paradigm to automatically register the referential frame of the intraoperative model with that of the endoscopic camera, without any external tracking system nor analysis of the endoscopic image. By including the distal part of the endoscope within the intraoperative acquisition field and holding it with an articulated arm, we are able to estimate the direction of the optical axis and the position of the optical center in the reconstructed volume. This approach allows us to determine directly the correspondence between the endoscopic camera and the 3D imaging system (Artis Zeego, Siemens). Obviously, this relationship holds only when the camera remains static, but this may occur several times during an intervention such as a liver segmentectomy. First, static AR can be used to reveal hidden vessels during the outlining of the resection planes by electrocoagulation. Second, AR can be used during the resection stage itself to identify encountered veins as hepatic or portal ones, as the former requires electric scissors and the latter clipping. Third, overlays can help guide the needle during a thermal ablation. Every time the endoscope is displaced, static AR can be updated by performing another intraoperative acquisition. If continuous endoscopic navigation is required, then we could resort to a classic tracking technique such as simultaneous localization and mapping (SLAM) or an external optical tracking apparatus, but our approach would still remove any need for calibration of this system. Finally, provided that a transfer function allows a good visualization of critical

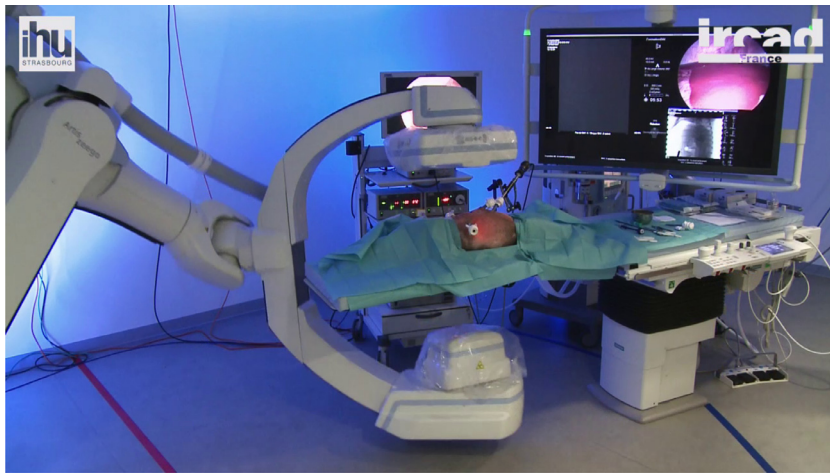


Fig. 2. Picture of a typical setup in the hybrid operating room.

structures in the volume rendering mode of the intraoperative images, we are then able to directly achieve AR without preoperative imaging.

In a sense, a common hand-eye calibration method seems to be applicable to our system. However, due to the absence of external tracking, the relationship between a calibration object and the camera can only be determined by means of an additional 3D CT acquisition. If performed before the operation, this acquisition may not be useful as the zoom, focus and position of the light source are not yet set as desired by the surgeon (proof that these parameters influence the augmentation is provided in Appendix D). If performed once the patient is on the table, a new acquisition is not possible without moving the patient. Either way, this would lengthen the surgical workflow more than with the proposed method.

After previously publishing a brief description of our method (Bernhardt et al., 2014), the present article presents further details and introduces several major supplementary contributions. For the intrinsic parameters, we assumed in our previous work the coincidence between the endoscope axis and the optical axis. Here we present a method to solve this issue. For the extrinsic parameters, we introduce a new 3D template-based technique for the endoscope extraction from the intraoperative acquisition. Concerning the quantitative results, we have created a new 3D-printed calibration chessboard with better accuracy and have performed significantly more experiments. For the qualitative results, we provide additional *in vivo* visuals, including new textured surface-based renderings.

The article is organized in two main sections. The first describes the system setup and details our method to determine the camera parameters. The second section provides quantitative and qualitative results from various experiments. Finally, a conclusion summarizes the main achievements, provides discussions and exposes some of our future plans.

2. Methodology

2.1. Description of the system and patient setup

This Section describes the typical intervention setup that we aim at with our method. The patient lies on the operation table near to a 3D rotational C-arm machine, as shown in Fig. 2.

After the pneumoperitoneum is performed, the surgeon inserts the endoscope inside the patient through a port and adjusts the zoom and focus of the camera as well as the position of the light source position (some endoscopes can rotate around the camera,

see Fig. 21 in Appendix D). In this work, we only consider monocular 0° angled endoscopes as they are still widely used for interventions and their optics is easy to model. Nonetheless, we discuss the adaptation of our method to 30° endoscopes and stereoscopes in Section 4. Once the camera settings are fixed, the endoscope is pulled out of the patient and two calibration steps are performed (as described in Sections 2.2 and 2.4). These involve only an analysis of the camera output and are typically achieved within a minute. The endoscope is then inserted back in the abdominal cavity and positioned so as to obtain an optimal view of the abdominal region of interest, but also so that it will appear in the intraoperative 3D image acquired by the C-arm (see Fig. 3). One can make sure it is sufficiently inserted by performing the common 2D fluoroscopic acquisitions that are parts of the clinical routine.

One may argue that the presence of the metallic endoscope would introduce artifacts in the CT reconstruction. Nonetheless, at a nominal distance from the organs, the endoscope produces only minimal artifacts. Most importantly, these are located underneath the endoscope tip i.e. outside the region of interest (ROI), as shown in Fig. 4. The contrary is unlikely, as the endoscope would have to be positioned right above the ROI, which is seldom in this kind of intervention.

To accurately augment the endoscopic view with intraoperative data, we must create a virtual camera with the same view upon the ROI as the actual endoscope. Thus, the virtual view will match the endoscopic image content. Therefore, we have to determine all the intrinsic and extrinsic parameters of the endoscopic camera.

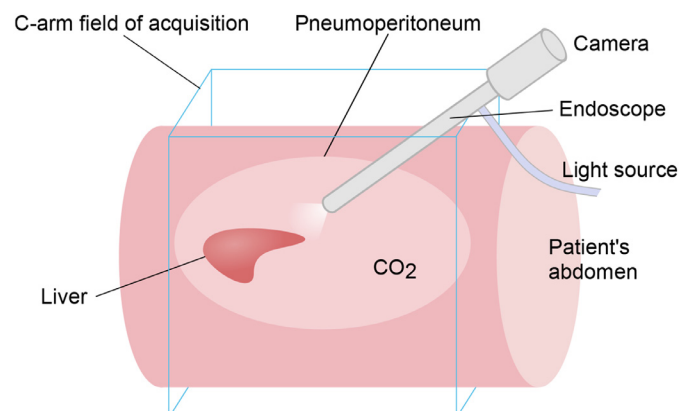


Fig. 3. Illustration of our typical abdominal intervention.

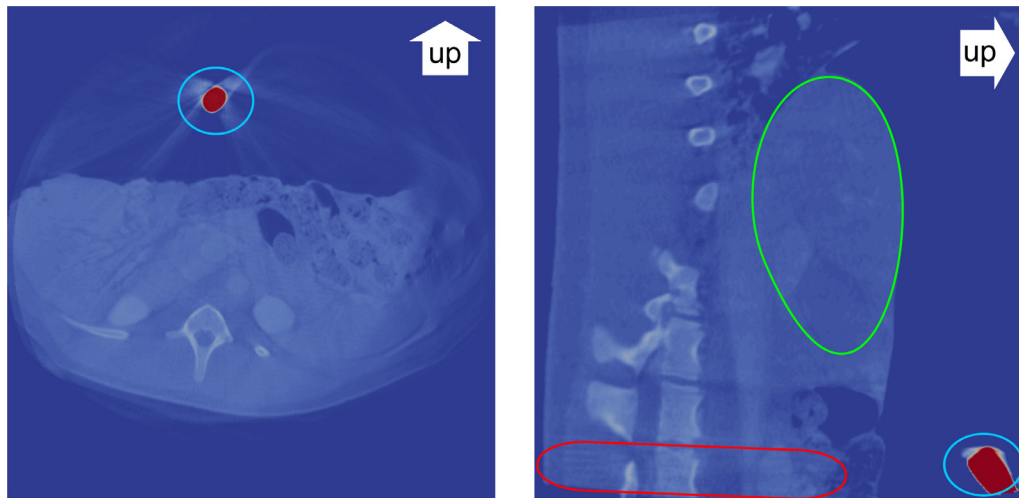


Fig. 4. Illustration of endoscope-induced artifacts in pig image. The axial view (left) shows that the endoscope (outlined in blue) produces thin artifacts in the reconstructed volume (zoom in for details). However, the sagittal view (right) reveals that these (outlined in red) are only located underneath the inserted tip (in blue) and do not degrade the region of interest which is the liver (in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

The intrinsic parameters are composed of the field of view, the position of the optical center and the distortion of the lenses. The extrinsic parameters consist in the location and orientation of the camera in the model space (i.e. the 3D image space). Finally, there is also an uncommon variable that comes into play in our method, which is the misalignment of the image sensor and the axis of revolution of the tube containing the optics. In the following Sections, details are given on how all these parameters are estimated.

2.2. Camera intrinsic parameters

After the surgeon has set the optics as desired, we capture a set of images of a generic chessboard pattern with several orientation angles and determine the intrinsic parameters through a classic camera calibration based on Zhang's method (Zhang, 2000). Thereby, we can notably estimate the focal length, from which the field of view (FOV) for the virtual camera can be calculated. Additionally, the calibration provides the position of the optical center in the image plane and the radial distortion of the lens, which can be used to undistort the endoscopic image. The quality of the calibration is commonly measured by the error from reprojection.

For our experiments, we typically got reprojection errors below 0.7 pixel.

2.3. Camera extrinsic parameters

To determine the extrinsic parameters, we use the information about the endoscope tip included in the volume image.

Camera pose estimation - The endoscope is mainly composed of titanium and thus is highly opaque to x-rays. Therefore, it yields very high values in the CT image, much larger than those of surrounding artifacts or any human tissue, as shown in Fig. 5. Extracting the endoscope then becomes trivial as we have a large range to set a threshold value for segmenting its voxels.

In our previous publication (Bernhardt et al., 2014), we described a method to estimate the endoscope position that was based on the analysis of the segmented voxels. Since then, experiments have shown that the accuracy of the registration was strongly dependent on the value of the selected threshold. In order to alleviate this issue and for the sake of providing a fully automatic AR approach whose results depend mostly on the data and not on the parameters, we developed a new method based on matching a 3D template.

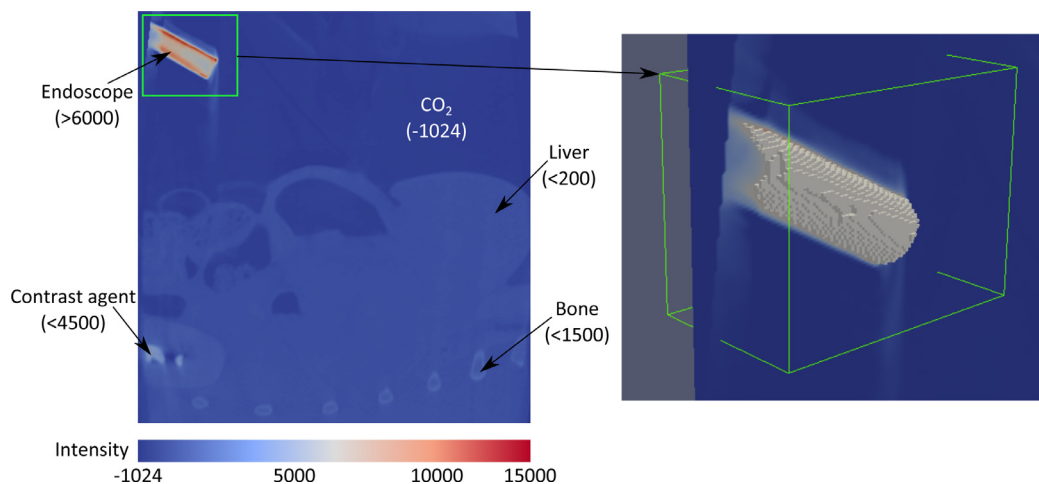


Fig. 5. The image on the left is a slice of the CT image, going through both the endoscope and a pig's anatomy. The intensity values are displayed for different regions of the scene, showing that the endoscope voxels exhibit very high contrast with any other part of the volume. Therefore, it is simple to extract the voxels corresponding to the endoscope (right image) and to reduce the search for its position only to its bounding box (green).

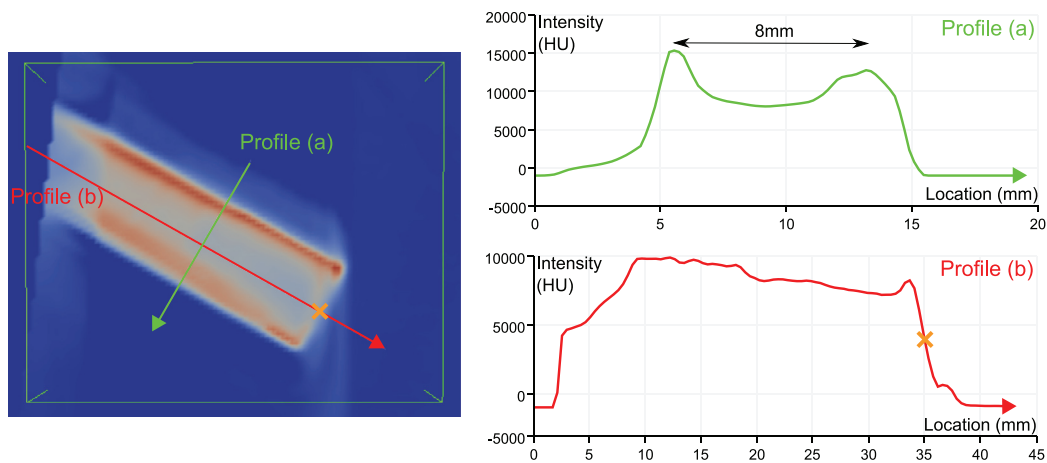


Fig. 6. The intensity profile (a) (in Hounsfield units or HU) across the endoscope section i.e. along the green line shows that the distance between the two edges corresponding to the surface of the endoscope may be less than its actual radius (10 mm). The intensity profile (b) along the red line shows the abrupt drop occurring at the tip location (orange cross). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

As one can see in the Fig. 6, the intensity peaks at the location of the surface of the endoscope. Therefore, knowing that it is tubular, we can retrieve the position of the endoscope by fitting a 3D cylindrical model on its surface. This is performed by maximizing the average intensity value sampled in the 3D CT data along this template (see Fig. 7). This sampling is obviously limited in space by the bounding box of the voxels previously extracted and the distance between two consecutive sampling points is related to the 3D CT image spacing ($0.49 \times 0.49 \times 0.49$ mm) so as to hit every voxel along the template at least once. As shown in Fig. 7, this model is represented with the position of a point C on its axis of revolution (coordinates along the X-, Y- and Z-axes), the two first Euler angles of its axis (ϕ and θ) and its radius R as the latter may differ from the actual endoscope width (see Fig. 6). These parameters are initialized with respect to the outputs of a principal component analysis (PCA) applied to the segmented voxels. By considering the average intensity value along the template as an objective function, matching is optimized using Powell's conjugate direction algorithm.

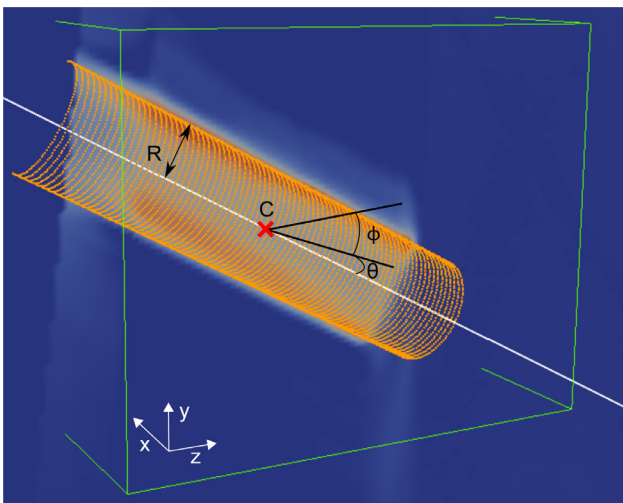


Fig. 7. This image shows the cylindrical template over a slice of the endoscope in the 3D CT data. The sampling points (orange) are positioned with respect to a point C on the axis of revolution, the Euler angles ϕ and θ and the radius R . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

The position of the endoscope tip (orange cross in Fig. 6) is found by thresholding a large negative gradient in the intensity profile along the determined axis as shown by the graph in Fig. 6(b) (blue stripe). It is worth noting that the resulting tip location will not actually correspond to the location of the virtual camera on the axis, notably because it depends on the zoom factor and the focus. Nonetheless, calibration experiments have shown that, for standard FOV and focus settings for such interventions, the correct location is very close to the actual tip of the endoscope and does not vary significantly across experiments. More details and experimental proof about this assertion are provided in Appendix A.

Also, given that we mainly worked with 0° endoscopes, there is a near perfect parallelism between the endoscope and its optical axis, as also demonstrated in Appendix A. Therefore, we safely set the orientation of the virtual camera to that of the endoscope. Discussions regarding non-parallelism for angled endoscopes and stereoscopes are included in Section 4.

So, we are able to determine all the extrinsic parameters but one; the tubular shape of the endoscope prevents the determination of the roll angle around the axis of revolution.

Camera roll estimation - This angle is automatically estimated by using a high-end 3-axis accelerometer designed for low accelerations (model LIS331DLH from ST Microelectronics) included in the camera. It is aligned and soldered on the printed circuit board (PCB) together with the imaging sensor, thus ensuring their alignment within the typical reflow assembly margins ($\pm 10 \mu\text{m}$, resulting in a negligible $\pm 0.2^\circ$ for this 3 mm-wide chip) (see Fig. 8). Moreover, this sensor is not considered as an external tracking device as it is already part of the endoscope.

So, the roll angle measured by the accelerometer is also that of the camera with respect to the gravity field. The interventional CT system is very precisely calibrated to provide a volume with its vertical axis aligned with the Earth's gravity. Thus, the roll angle of the virtual camera with respect to the 3D CT data can be set to the value given by the accelerometer. The accuracy obviously decreases as the endoscope moves towards verticality. A cone of about 10° around the vertical axis is typically a satisfactory boundary. Although, given our setup (see Fig. 3), the endoscope is never placed vertically as this would prevent the interventional C-arm from rotating around the surgical site. Also, the manufacturer reports a variation of sensitivity with respect to the tilt angle outside of verticality. However, thanks to trigonometric formulae and a combination of the acceleration values along the three axes, the sensitivity can become independent from the actual tilt angle

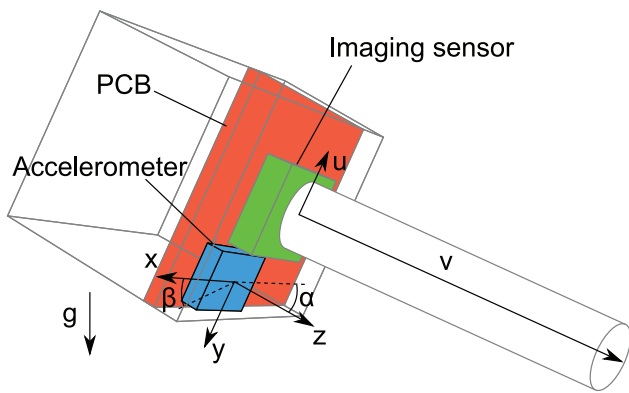


Fig. 8. Simplified illustration of our endoscope and its sensing components. The scene is projected along the optical axis v onto the imaging sensor (in green) and forms an image whose up vector is u . The CCD chip is soldered on a PCB (in red) together with the accelerometer (in blue) which measures the tilt angles with respect to gravity denoted g . Therefore, its roll angle β and the pitch angle α are the same as those of the imaging sensor. The dashed lines represent horizontality. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

values. Therefore, the roll angle estimation remains accurate in all cases.

We performed a quantitative evaluation of the roll angle estimation by comparing the value provided by the accelerometer and the one found by manually setting the rotation while trying to match visual cues in the scene. Across 14 experiments, we obtained an error of $0.15^\circ \pm 0.24$, which is larger than the value of 0.1° reported by the manufacturer, but still sufficiently accurate for our purpose.

2.4. Determining misalignment between endoscope tube and CCD sensor

Most of cameras, laparoscopic endoscopes included, present a small misalignment between their optical axis and the center of their imaging sensor. This offset between the principal point and the image center is usually estimated by a camera calibration and

is part of the intrinsic parameters. Additionally, in our case, it is important to note that the optical axis may differ from that of the actual tubular endoscope (see Fig. 9). As a result, the center of the imaging sensor is unlikely to coincide with the axis of revolution of the tube.

Since we position our virtual camera with respect to the shape of the endoscope (more specifically on its axis of revolution), we need to determine its geometrical relationship with the image plane as even a few micrometers of misalignment would result in several pixels of registration error. More specifically, we place the virtual camera onto the axis of revolution Δ of the endoscope, hence we ought to locate the projection C_Δ of Δ onto the CCD sensor. Then, in the endoscopic image of center C , since $\overrightarrow{CC_\Delta}$ is measured in pixels, we can compensate for the misalignment by translating the virtual image by $\overrightarrow{C_\Delta C}$ in the final registration.

The virtual and actual cameras having different origins (the former is located on the endoscope axis and the latter on the optical axis), $\overrightarrow{C_\Delta C}$ can also be interpreted as a 2D correction of a perspective effect. However, the discrepancy between the endoscope and optical axes is very small in both distance (a few millimeters at most) and parallelism (see Appendix A) compared to the dimensions of the scene. Thus, the orthographic projection of the virtual camera is very similar to the perspective of the actual camera. Also, since this correction is operated from the center of the image, the approximation is all the more accurate in this area, which is where the region of interest usually lies.

Calibration with a square tube - Provided that the rod is a perfect cylinder, all the diameters of the endoscope tube intersect on its axis of revolution. Therefore, a practical way of estimating Δ is to determine several of these diameters in the endoscopic image and calculate their intersection. A simple technique to do so is to record images and extract the corners of a square-profiled tube resting on top of the endoscope tip, as illustrated by Fig. 10. The actual method to estimate the diagonals is described in Appendix B.

So, while the square tube slowly rotates on the endoscope, we extract the diagonals in the images and associate them from one frame to the next simply by considering their orientation similarity. This results in two sets of intersecting diagonals (see Fig. 11).

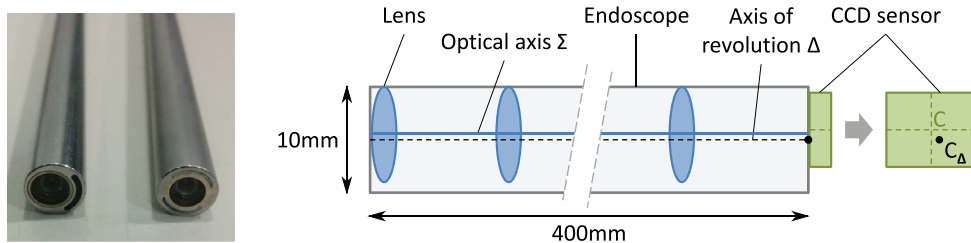


Fig. 9. At the top is a picture of two endoscope tips. The left clearly has its optics decentered with respect to the tube because of the light emitter. The right seems to be centered, but could still result in several pixels of registration error. At the bottom is a simplified illustration of the structure of a rigid endoscope with exaggerated misalignments. One can see that the center C of the imaging sensor does not necessarily coincide with the optical axis Σ which, in turn, may very well not be aligned with the axis of revolution Δ of the rod. As a result, the projection C_Δ of the endoscope axis may diverge from the image center C , resulting in an registration error in translation.



Fig. 10. A square-profiled tube rests on the distal part of the endoscope. The line passing by the top and bottom corners (red) intersects the sought axis of revolution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

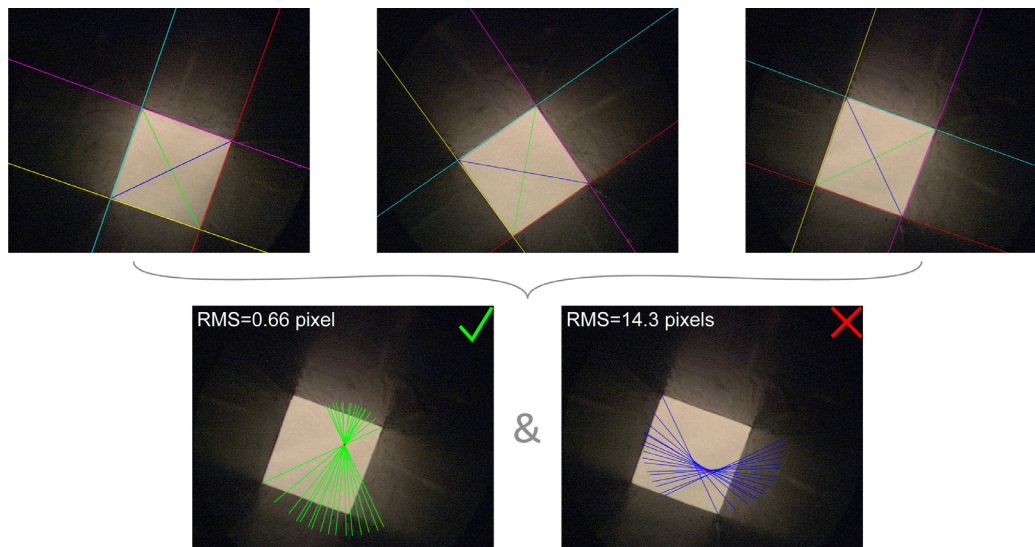


Fig. 11. At the top, the edges and thus the diagonals of the squares are detected in a sequence of images, resulting in two sets of lines, the green ones and the blue ones, as illustrated at the bottom. Using a least squares method, we can calculate, for each set, the point that minimizes its distance from the lines. The set with the smallest discrepancy (here the left one) is the closest to a perfect intersection of all the lines at the same point, which is considered to be C_{Δ} . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

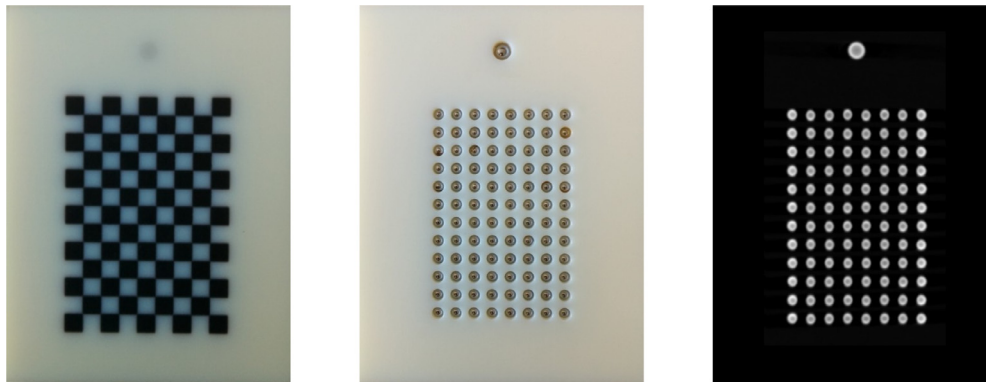


Fig. 12. On the left is a front picture of the 3D-printed chessboard. At the middle, a back picture shows the 3 mm steel balls inserted underneath the corners. On the right, the same balls seen from the CT image. On top of the chessboard, a bigger 5 mm ball is inserted, observable from both CT and endoscope, which resolves ambiguity in the orientation.

To determine which set corresponds to the sought projection, we calculate the position of the point that minimizes its distance to all lines for each set. In the correct case, all the lines should intersect roughly at the same point and therefore the deviation should be the smallest.

Using a least-square formulation, we calculate the intersection of the diagonals as the point that minimizes its distance to each line (equations are provided in Appendix C). As a result, we are also able to determine and compensate for the misalignment between the endoscope axis and the CCD sensor. Thus, the endoscopic view can be augmented with an accurately registered virtual image of the model. In Appendix D, we provide quantitative results about how the determination of C_{Δ} is influenced by the insertion depth in the square tube, the zoom, the focus and the position of the light source. In the next section, we present the results from numerous experiments that we have performed to assess the accuracy of our method.

3. Experimental results

Various experiments have been performed, including quantitative ones to evaluate the accuracy and qualitative studies show clinical potential.

3.1. Quantitative results and validation

We carried out a significant number of experiments to validate quantitatively the accuracy of our method.

Protocol - A dedicated chessboard has been created to assess our method's accuracy (see Fig. 12). It has been printed in 3D in order to achieve a 0.1 mm accuracy for the pattern. Composed of 13×9 squares of 5×5 mm, it fills most of the field of view of an endoscopic camera at the nominal distance. On one side, the regular chessboard pattern is used by the camera to locate precisely each corner. On the other side, right underneath each corner, a 3 mm steel ball is inserted at a known depth of 3.8 mm. Since steel also highly attenuates x-rays, we are able to segment each of them and thus calculate the position of each corner in the volume image too. Hence, by placing the endoscope so that the chessboard fills most of its view, it is possible to evaluate the precision of the registration over most of the camera field of view.

Following the protocol described in Section 2.1, once the camera has been tuned as desired and calibrated, the chessboard is acquired by both modes of imaging (3D C-arm and endoscope). Using our method, we locate the endoscope in the CT volume and both roll angle and potential misalignment between the endoscopic rod and the CCD sensor are determined. Then, we independently

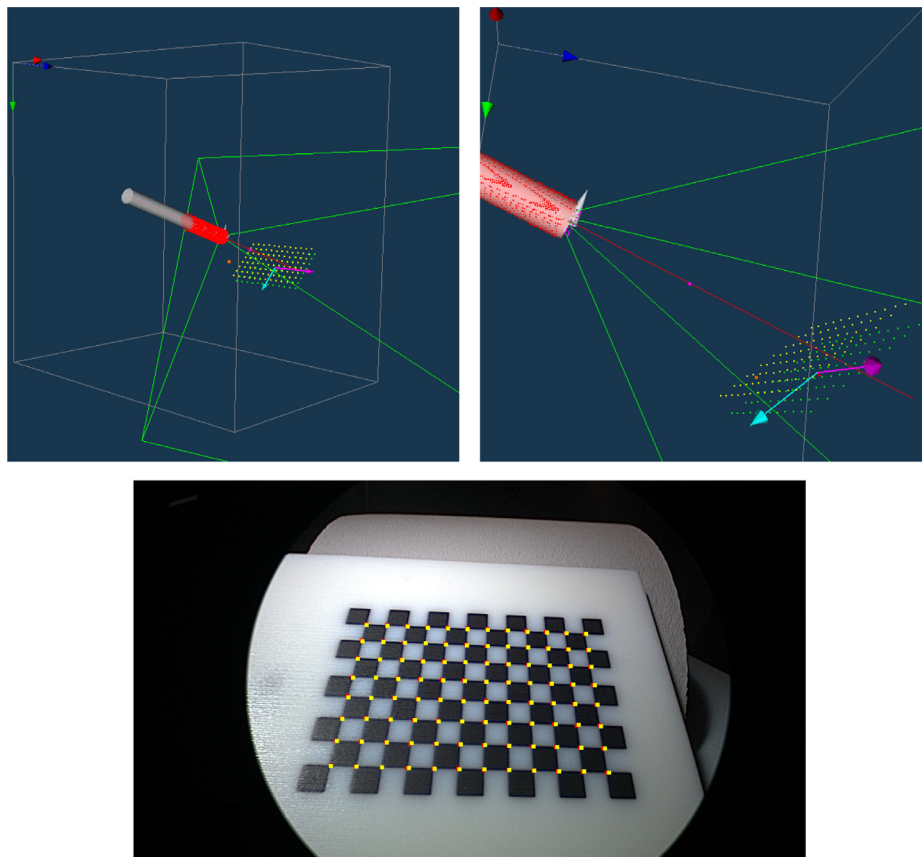


Fig. 13. On the left are two views of a rendering of the scene with the outline of the C-arm volume, the virtual camera placed according to our calculations from the segmented voxels (red dots) and its frustum (green). The green dots are the centers of the segmented steel balls, which we use to infer the position of the actual chessboard corners in yellow by a known translation. On the right is the endoscopic image with the corners located in red. The yellow points from the virtual view are projected in the same image for comparison (zoom in for details). For this experiment, we obtained $\epsilon = 6.38 \pm 2.03p$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

determine the chessboard corner locations from both the CT volume and the endoscopic image with respectively ITK (threshold, labeling) and OpenCV (subpixel corner extraction). By performing the registration, the corner locations from the volume rendering are projected in the 2D space of the endoscopic image. Therefore, we are able to calculate the registration error ϵ in pixels between the two sets of points across the 1080p image, as shown in Fig. 13.

Results - We performed this experiment 28 times with five endoscopes of the same model (Karl Storz Hopkins II) but of different generations. Those of the first generation are decentered and those

of the second generation are not (respectively left and right in the picture in Fig. 9). Additionally, for each experiment, we used different camera settings (zoom, focus, etc.) and placements for both the chessboard and the camera. Moreover, as the endoscopic tube and the camera are separable for straight endoscopes, they can rotate around each other, constituting yet another variable across experiments. Finally, due to their use and manufacturing, the optical properties and shape of these endoscopes may slightly differ from one to the other. All the results about the residual error (average and standard deviation) are reported in Fig. 14.

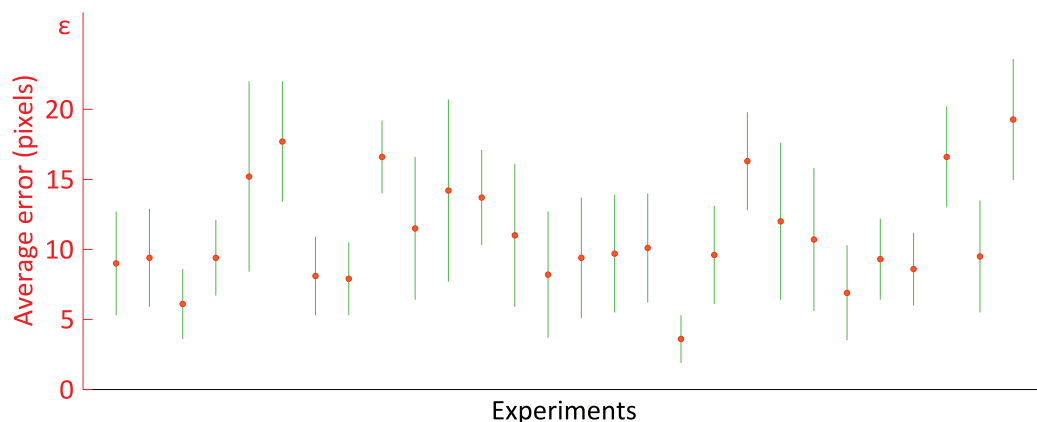


Fig. 14. This graph shows the results for all 28 experiments. The average registration error across all 96 chessboard corners is displayed in red and the corresponding standard deviation in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

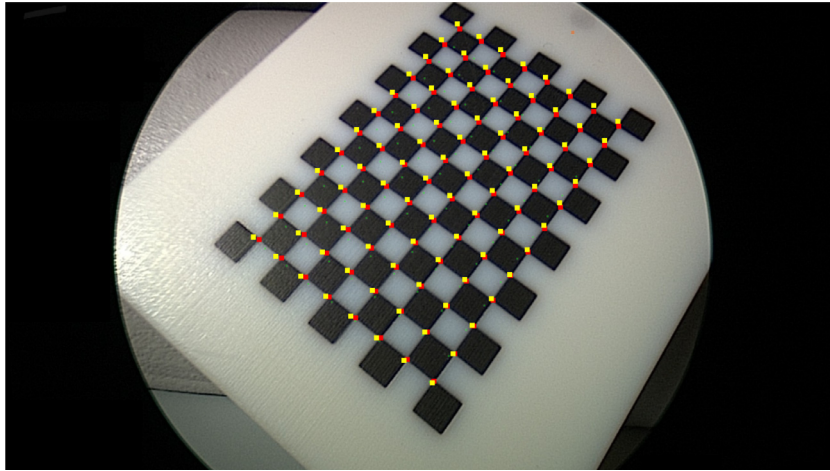


Fig. 15. This is the worst registration we obtained across the 28 experiments with 1080p images.

These show that the registration error is consistently low for the 28 experiments. The largest registration error is smaller than 24 pixels in a 1080p image and most of the experiments present an average error smaller than 15 pixels, which approximately correspond to a millimeter at the nominal distance. In order to facilitate the appreciation of the results, we show the worst registration that we obtained in Fig. 15.

According to the surgeons at our institution and current standards for intra-abdominal interventions, these errors are sufficiently small to fulfill AR purposes for surgical procedures. A typical run-time we measured on a computer with an Intel i5 processor at 3.30 GHz was 7 s for the camera calibration over 9 images, 25 s for the square tube calibration over 14 images and 6 s for the other processing, including rendering. Besides, the online tasks are the acquisition of the checkerboard for the camera calibration (15–30 s) and the acquisition of the rotating square tube (10–20 s). The only offline task is to make sure that one is able to set a correct threshold to extract the endoscope in the CT volume. To do so, only one acquisition of the endoscope with the C-arm is required.

Error sources analysis - We have measured the relationship between the registration error and the depth at which the endoscope is inserted into the volume, in order to provide some kind of recommended minimum insertion guideline. The results are shown in Fig. 16 and were performed on 12 different endoscope positions.

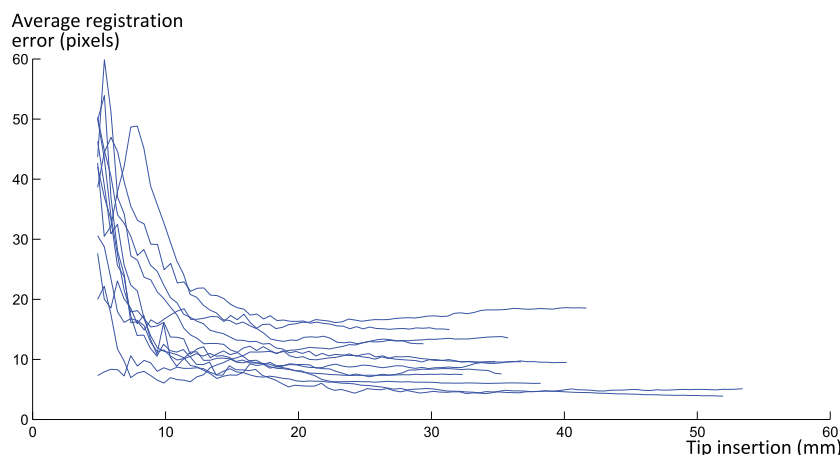


Fig. 16. For each of 12 experiments, the average registration error in pixels is displayed with respect to the tip insertion depth in millimeters at which the endoscope is inserted in the volume image.

One can see that all registration errors stabilize beyond 17 mm, which then becomes our recommendation for the tip insertion depth to ensure the accuracy of our method.

However, even with the endoscope well inserted inside the volume of acquisition, the registration error is still influenced by other parts of the process. To assess how much each part affects the final output, we performed ten distinctive times each step of the complete process (camera and square tube calibrations, 3D CT and endoscopic images) for three different setups. Each setup consists of a different pose of the endoscope and the radio-opaque chessboard. Moreover, the acquisitions have been performed with as much variation as possible within each setup by changing the lighting of the scene and choosing various chessboards for the camera calibration. Table 1 reports the average error and the standard deviation of registration in pixels measured when only one step of the process changes. This concerns the square tube calibration, the camera calibration, the endoscopic image (corners localization) and the 3D CT image (endoscope extraction and corners localization).

From the standard deviations reported in Table 1, one can notice that the camera calibration does not significantly affect the registration accuracy, probably due to the reliability of Zhang's method for the endoscopic camera calibration and the robustness of the corner detection method. The latter may also explain why variations in the endoscopic image does not impact the registration. However, successive acquisitions of a same scene by the 3D

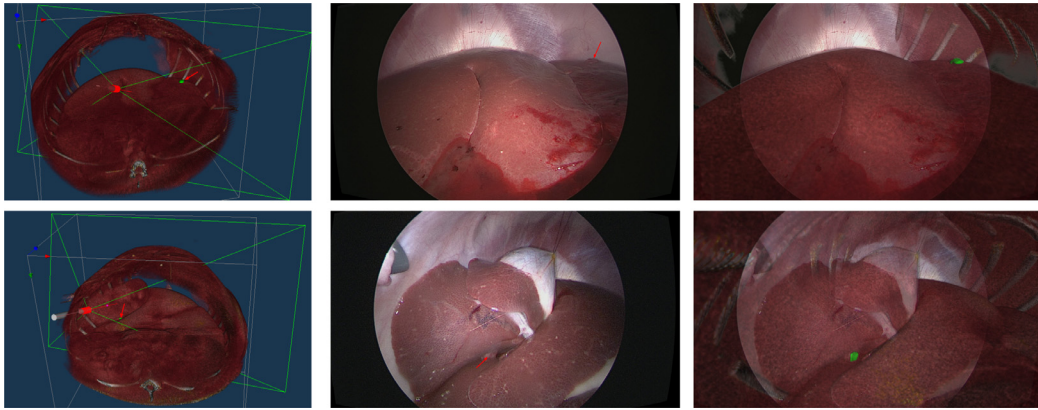


Fig. 17. The volume rendering (left) is projected onto the endoscopic image of the liver (center) to augment it (right). For each case, a tumor has been manually segmented from the intraoperative volume image and is pointed by the red arrow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

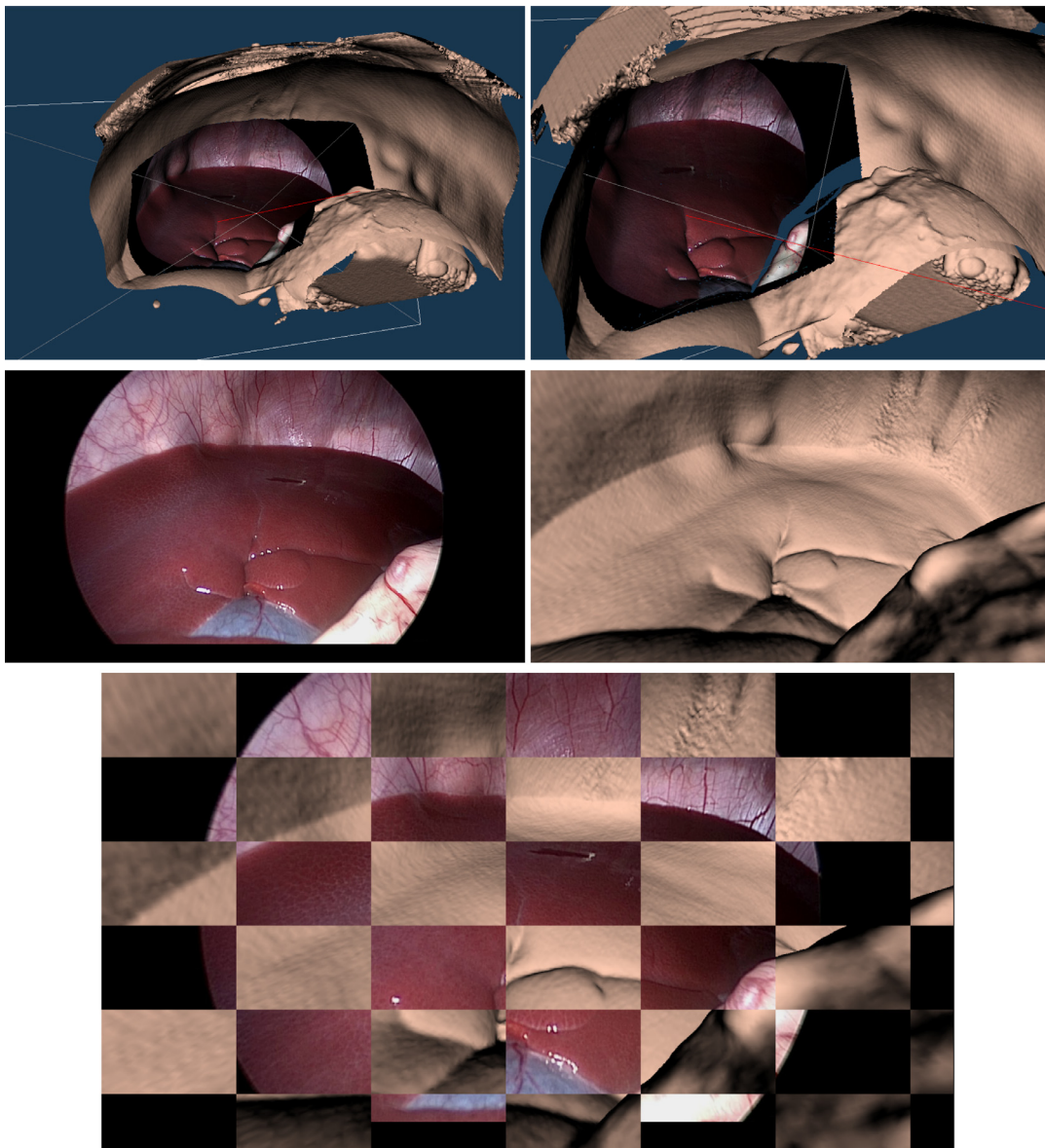


Fig. 18. At the top, the endoscopic image is projected onto the surface of the abdominal cavity from a virtual camera placed according to our method. One can notice on the top right how well-aligned the edges are in both the image and the mesh. In the middle are the endoscopic image (left) and the view of the surface from the virtual camera (right), both showing the same scene. Their good correspondence is shown by a grid mosaicing of these two images (bottom).

Table 1

Average and standard deviation of the registration error (in pixels) observed when only one part of the process changes ten times and across three different setups. While one step changes, all the others are assigned the values corresponding to the best result in terms of accuracy.

Experiment setup	1	2	3
Square tube calibration	16.25 ± 7.51	15.11 ± 6.23	13.34 ± 7.16
Camera calibration	13.93 ± 0.28	11.01 ± 1.01	6.43 ± 0.36
Endoscopic image	13.81 ± 0.07	11.76 ± 0.09	6.75 ± 0.11
3D CT image	14.76 ± 0.79	14.80 ± 1.84	6.76 ± 2.39

CT imaging system may yield non-negligible fluctuations in the intensity of the voxels. Since our method relies on voxel intensity to estimate the endoscope pose in the volume image, we also logically observe that this part of the process affects the final output of our method. Nonetheless, so far, these variations represent at most only tenths of a millimeter in the augmented scene and may be ignored. On the contrary, various square tube calibrations may change the augmentation by up to 14 pixels which corresponds to almost a millimeter in the scene. Therefore, this routine has by far the most impact on the registration accuracy and ought to be performed with caution. If done so, one can then obtain precise augmentations like the ones performed on *in vivo* data and presented in the next Section.

3.2. Qualitative results with interventional data on pig

So as to assert the clinical feasibility of our method, we performed five sessions of acquisition on pigs. For each of them, we performed a camera calibration and estimated its position with respect to the scene using our method, enabling us to augment the endoscopic image with the volume rendering. Fig. 17 shows instances of augmentation of the liver surface of two different pigs with lesions. The lesions are manually segmented from the intraoperative images and displayed through the registration using our method. One can see that the liver contour and the lesions from the volume image are well-aligned with their representations in the endoscopic view with a registration error smaller than 10 pixels which corresponds to less than a millimeter at the nominal distance.

3.3. Qualitative results with surface rendering

As a secondary way of visually validating our registration method, we also implemented a way of projecting the endoscopic image back to the C-arm volume data. By applying the marching cubes algorithm on the volume data, we are able to extract the surface of the pneumoperitoneum. Then, by placing a virtual camera using our method, the image can be back-projected as a texture onto the visible cells (see Fig. 18). One can assert the registration accuracy by observing how well the contours on the surface, visible due to shadowing, match their counterparts in the endoscopic image. Beside validation, this process can also be used in order to automatically produce highly photo-realistic textured 3D models, hence ridding users of manually identifying the correspondences between the surface and the endoscopic image. Such models may prove useful for education as well as for surgery simulation.

4. Conclusion and discussion

In this article, we have presented a new simple paradigm to register an intraoperative model with the endoscopic image. By including and analyzing the endoscope tip inside the 3D C-arm field of acquisition, we can precisely determine almost all the camera

extrinsic parameters, thus allowing static AR with no need for any external tracking device.

During quantitative experiments, our method has proven to be robust and accurate throughout many evaluations using a dedicated calibration object, with most registration errors below 15 pixels (i.e. approximately 1 mm in the scene at nominal distance). Regarding the ground truth accuracy, it could be argued that we extract the steel balls from the volume image with a fixed threshold, which may reduce the accuracy. Tuning its value yields slight changes in the position of the balls which in turn displaces the inferred chessboard corners in the volume rendering. Nonetheless, since these uncertainties are of less than a fraction of a millimeter, they do not significantly impact the ground truth accuracy, even combined with the imprecision of the 3D printing (0.1 mm).

In addition, as mentioned in Section 2.1, all experiments were performed so far with straight endoscopes as most of laparoscopic interventions are performed with these. Nonetheless, we are aware that 30° optics is preferred in some specialties such as thoracic and bariatric surgery, but by definition these devices do not ensure parallelism of their optical and shaft axes. Some initial tests performed with such an endoscope and an estimation of the bevel direction from the 3D CT data did not give satisfactory results since the distal tip is composed of too few voxels. Nonetheless, we argue that it is very likely that an improved version of our square tube, combined with the accelerometer output, would allow an accurate estimation of the 30° angle direction with respect to the camera. Likewise, we also performed a few tests with a stereoscope, which has all its optical components concentrated at the tip and thus also does not ensure parallelism between the optical and the endoscope axes. But the results were promising and we are convinced that an adaptation of the square tube calibration would allow the use of our method with such endoscopes. In other words, we believe in the genericity of our approach and plan on making it accessible for any kind of endoscope.

In conclusion, we highlight the fact that this is the first AR method in laparoscopy on *in vivo* data to attain a accuracy of 1 mm at nominal distance, contrary to methods based on optical tracking, whose accuracy is hindered by all the calibration process. Therefore, as shown by our *in vivo* experiments, our AR system can be used to automatically augment the endoscopic view with useful information from the 3D intraoperative imaging system (e.g. volume rendering or segmentations).

Currently, the image quality of intraoperative 3D C-arms may not be sufficient to easily extract features of interest, but a huge amount of work is being carried out by manufacturers to significantly improve intraoperative image quality. Meanwhile, our approach may also reveal itself to be advantageous for static AR from preoperative data if its relationship with the intraoperative image is already determined. In case of endoscope displacement, a new acquisition can be made to update the AR. One concern may then be the successive radiation exposures of the patient, but these are already three times less than with regular CT scanners and keep on improving. Thus, it is reasonable to assert that hybrid operating rooms will become increasingly popular.

If true dynamic AR is required, we could resort to an image analysis technique such as SLAM or an external system such as optical tracking or a robotic arm, but our method would still remove any need for calibration, making it potentially more accurate. Moreover, the accelerometer also provides useful motion data that could be used to refine the tracking. Indeed, an accelerometer is accurate for fast motions, making it complementary to an optical tracking system, which is more tailored for slow movements. We also plan on using a passive robotic arm.

Finally, we are in the process of optimizing and integrating our algorithms so that they can work seamlessly in an operating room; the first tests on patient are expected in 2016.

Acknowledgment

The authors would like to thank the University of Strasbourg, the IHU of Strasbourg for funding this project, the IHU imaging technicians for making the experiments possible, the IRCAD Strasbourg R&D development team, Johannes A. Fallert from Karl Storz GmbH, Tuttlingen, Germany and Philip Mewes from Siemens Healthcare GmbH, Munich, Germany for providing useful information about their equipments.

Appendix A. Camera position and parallelism with respect to the endoscope

To determine the position of the camera from the endoscope tip, we typically perform an optimization of the position of the virtual camera along the endoscope axis for several experiments on the radio-opaque chessboard (see Section 3.1). It has been observed that this distance d is approximately 3.4 ± 2 mm for standard FOV and focus settings. Also, this deviation of ± 2 mm introduces only between 1 to 7 pixels of reprojection error depending on the perspective of the scene. For the sake of consistency, we also calculated the position of the camera from its view of the chessboard by solving the classic Perspective-n-Point (PnP) problem and measured its distance from the tip along the endoscope axis (see Fig. 19). Across 28 experiments, we got $d = 4.2 \pm 0.43$ mm, which is close to what we previously found. Therefore, this distance d can be set as a constant.

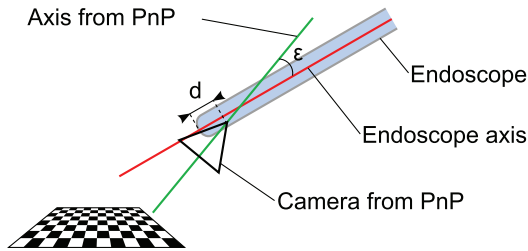


Fig. 19. Endoscope looking at chessboard. By detecting the corners and solving the PnP problem, we can evaluate the position and orientation of the camera and compare them to what we find with our method.

PnP solving also provides the camera orientation, which is helpful to estimate how parallel the optical and endoscope axes are in the case of a 0° endoscope. Across the same 28 experiments, we measured $\epsilon = 0.42 \pm 0.17^\circ$ (see Fig. 19), which introduces a registration error of approximately 0.51 ± 0.21 mm at nominal distance (70 mm). Therefore, it is safe to assume that the endoscope and the optical axes are parallel.

Appendix B. Square tube corner detection

To accurately determine the corners, the inside of the square tube is painted in matte black, so that the images obtained are very easily segmentable (see Fig. 20). First, a bilateral filter is applied to the image to reduce the noise while preserving the edges (Fig. 20(b)). Second, we use the Canny edge detector to produce the binary edge map of the image (Fig. 20(c)). Third, we apply the probabilistic Hough transform to extract the strongest segments from the map and cluster them by similarity of direction and location (Fig. 20(d)). Finally, we average each cluster of segments to obtain the square edges and thus the corners and diagonals (Fig. 20(e)).

Appendix C. Determination of the intersection of diagonals

Let us consider a set of N seemingly intersecting straight lines, each defined by a point P_i and a (normalized) leading vector v_i . The squared distance of a point P to the i th line is thereby:

$$d_i^2 = |P - P_i|^2 - ((P - P_i) \cdot v_i)^2 \quad (1)$$

If we consider the vector derivatives of Eq. 1, we get:

$$\begin{aligned} \frac{\partial d_i^2}{\partial P} &= 2(P - P_i - ((P - P_i) \cdot v_i)v_i) \\ &= 2(I - v_i v_i^T)(P - P_i) \end{aligned} \quad (2)$$

The point C_Δ that best approximates the intersection according to the least-squares sense is also the null space of Eq. 2 for every line i :

$$\sum_i^N (I - v_i v_i^T)(C_\Delta - P_i) = 0 \quad (3)$$

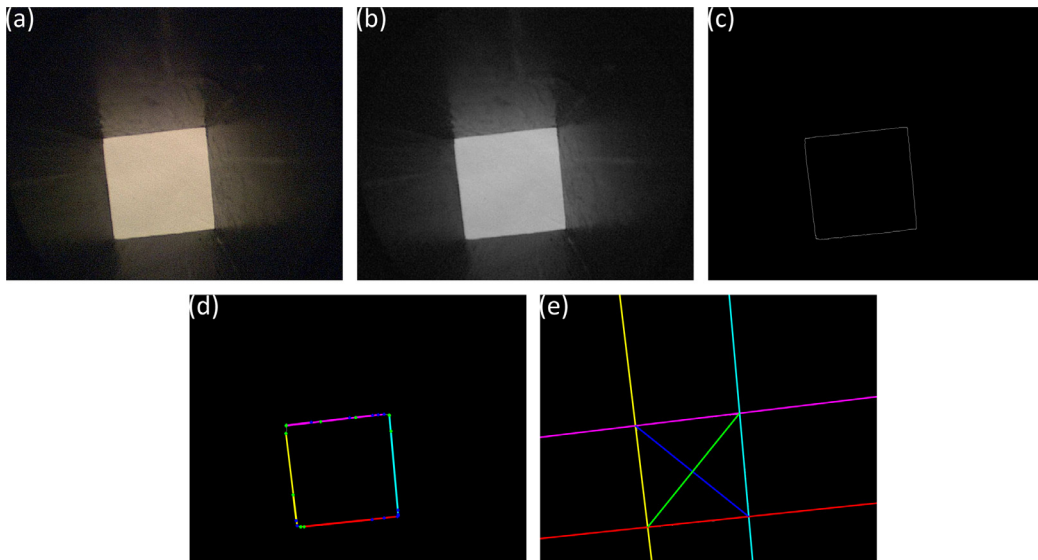


Fig. 20. Our method of square edge extraction: (a) Original image; (b) Bilateral filtering; (c) Canny edge detection; (d) Probabilistic Hough transform and separation of segments into four clusters; (e) Averaging of segments to obtain the edges and then the diagonals.

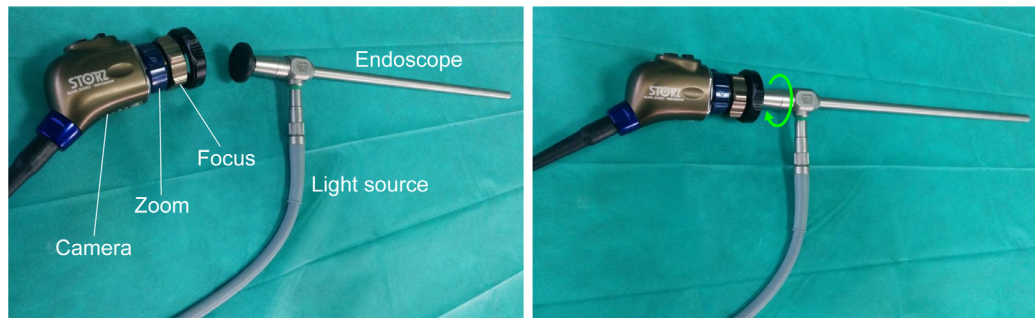


Fig. 21. Pictures of a separable endoscope. The endoscope and the camera can be separated (left) and attached (right) through a joint that allows a spinning rotation of the endoscope (green arrow). The light source can thus be placed according to the surgeon's desire during the intervention. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Therefore, we can obtain C_{Δ} from Eq. 3 as:

$$C_{\Delta} = \left[\sum_i^N (I - v_i v_i^T) \right]^{\dagger} \sum_i^N (I - v_i v_i^T) P_i \quad (4)$$

Appendix D. Influence of insertion depth, zoom, focus and light source position

Concerning the square tube calibration, one may wonder by how much should the endoscope be inserted in the tube. Since the endoscope is calibrated to capture the scene at a certain distance, it is naturally inserted so that the square end of the tube appears sharp in the image. Thus, the output plane is at that same nominal distance as the scene to be imaged and the 2D correction of the perspective effect is accurate. Nonetheless, for the sake of robustness, we measured the variations of C_{Δ} at ten different insertion depths around the depth of focus and we obtained a standard deviation of (0.9, 0.9) pixels along the image X- and Y-axes, which is negligible. This may also be considered as a proof that the perspective difference between the virtual and the actual camera is small.

In the introduction, we asserted that no calibration of the endoscope can be performed prior to the intervention as the zoom, focus and the light source position have an impact on the registration and may be changed by the surgeon (see Fig. 21). Therefore, we also measured the variations of C_{Δ} with respect to these parameters. For 11 different values of focus around a standard setting, the standard deviation of C_{Δ} is (2.0, 0.9) pixels along the image X- and Y-axes. For 16 different values of zoom between a medium zoom and a complete zoom-out, the standard deviation of C_{Δ} is (4.3, 7.5) pixels along the image X- and Y-axes. For 11 different positions of the light source, the standard deviation of C_{Δ} is (16.1, 21.1) pixels along the image X- and Y-axes. Hence, while changing the camera focus has little impact on the 2D correction, variations of the zoom and the light source position significantly influence the square tube calibration and thereby the augmentation.

References

- Bano, J., Hostettler, A., Nicolau, S., Cotin, S., Doignon, C., Wu, H.S., Huang, M.H., Soler, L., Marescaux, J., 2012. Simulation of pneumoperitoneum for laparoscopic surgery planning. In: Proceedings of MICCAI, pp. 91–98.
- Bano, J., Nicolau, S.A., Hostettler, A., Doignon, C., Marescaux, J., Soler, L., 2013. Registration of preoperative liver model for laparoscopic surgery from intraoperative 3D acquisition. In: Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions, pp. 201–210.
- Baumhauer, M., Feuerstein, M., Meinzer, H.-P., Rassweiler, J., 2008. Navigation in endoscopic soft tissue surgery: perspectives and limitations. *J. Endourol. / Endourol. Soc.* 22 (4), 751–766.
- Bernhardt, S., Nicolau, S., Agnus, V., Soler, L., Doignon, C., Marescaux, J., 2014. Automatic detection of endoscope in intraoperative CT image: application to ar guidance in laparoscopic surgery. In: Proceedings of ISBI. IEEE, pp. 563–567.
- Feuerstein, M., Mussack, T., Heining, S.M., Navab, N., 2008. Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. *IEEE Trans. Med. Imaging* 27 (3), 355–369.
- Figl, M., Rueckert, D., Hawkes, D., Casula, R., Hu, M., Pedro, O., Zhang, D.P., Penney, G., Bello, F., Edwards, P., 2010. Image guidance for robotic minimally invasive coronary artery bypass. *Comput. Med. Imaging Graph.* 34 (1), 61–68.
- Garcia, J., Malik, P., Liu, J., Kowal, J., Marti, G., Pappas, I., Caversaccio, M., 2004. Image-guided surgical microscope with mounted mini-tracker. In: International Congress Series, vol. 1268, p. 1311.
- Haouchine, N., Dequidt, J., Peterlik, I., Kerrien, E., Berger, M.-O., Cotin, S., 2013. Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery. In: Proceedings of ISMAR, pp. 199–208.
- Hu, M., Penney, G.P., Rueckert, D., Edwards, P.J., Bello, F., Casula, R., Figl, M., Hawkes, D.J., 2009. Non-rigid reconstruction of the beating heart surface for minimally invasive cardiac surgery. In: Proceedings of MICCAI, pp. 34–42.
- Kitasaka, T., Mori, K., Hayashi, Y., Suenaga, Y., Hashizume, M., Toriwaki, J., 2004. Virtual pneumoperitoneum for generating virtual laparoscopic views based on volumetric deformation. In: Proceedings of MICCAI, pp. 559–567.
- Lo, B., Chung, A.J., Stoyanov, D., Mylonas, G., Yang, G.-Z., 2008. Real-time intraoperative 3D tissue deformation recovery. In: Biomedical Imaging: From Nano to Macro, ISBI, pp. 1387–1390.
- Mahmoud, N., Nicolau, S., Keshk, A., Ahmad, M., Soler, L., Marescaux, J., 2012. Fast 3D Structure from motion with missing points from registration of partial reconstructions. In: Proceedings of International Conference on Articulated Motion and Deformable Objects (AMDO'12) LNCS 7378, pp. 173–183.
- Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., Kolb, A., Rodrigues, M., Sorger, J., Speidel, S., Stoyanov, D., 2013. Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery. *Med. Image Anal. J.* 17 (8), 974–996.
- Malti, A., Bartoli, A., 2014. Combining conformal deformation and the cook-torrance model for 3D reconstruction in laparoscopy. *IEEE Trans. Biomed. Eng.* 61 (6), 1684–1692.
- Marescaux, J., Rubino, F., Arenas, M., Mutter, D., Soler, L., 2004. Augmented-reality-assisted laparoscopic adrenalectomy. *J. Am. Med. Assoc.* 292 (18), 2211–2215.
- Mårvik, R., Langø, T., Tangen, G.A., Andersen, J.O., Kaspersen, J.H., Ystgaard, B., Sjolie, E., Fougner, R., Fjøsne, H.E., Nagelhus Hernes, T.A., 2004. Laparoscopic navigation pointer for three-dimensional image-guided surgery. *Surg. Endosc.* 18 (8), 1242–1248.
- Mountney, P., Yang, G.-Z., 2010. Motion compensated slam for image guided surgery. In: Proceedings of MICCAI, pp. 496–504.
- Nicolau, S., Soler, L., Mutter, D., Marescaux, J., 2011. Augmented reality in laparoscopic surgical oncology. *Surg. Oncol.* 20 (3), 189–201.
- Nicolau, S.A., Diana, M., Agnus, V., Soler, L., Marescaux, J., 2013. Semi-automated augmented reality for laparoscopic surgery: first in-vivo evaluation. *Int. J. Comput. Assist. Radiol. Surg.* 8 (1 suppl), 109–110.
- Nicolau, S.A., 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, pp. 124–133.
- Oktay, O., Zhang, L., Mansi, T., Mountney, P., Mewes, P., Nicolau, S., Soler, L., Chedhdhotel, C., 2013. Biomechanically driven registration of pre- to intra-operative 3D images for laparoscopic surgery. In: Proceedings of MICCAI, pp. 1–9.
- Pizarro, D., Bartoli, A., 2012. Feature-based deformable surface detection with self-occlusion reasoning. *Int. J. Comput. Vis.* 97 (1), 54–70.
- Sánchez-Margallo, F.M., Moyano-Cuevas, J.L., Latorre, R., Maestre, J., Correa, L., Pagador, J.B., Sánchez-Peralta, L.F., Sánchez-Margallo, J.A., Usón-Gargallo, J., 2011. Anatomical changes due to pneumoperitoneum analyzed by MRI: an experimental study in pigs. *Surg. Radiol. Anat. : SRA* 33 (5), 389–396.
- Shahidi, R., Bax, M.R., Maurer Jr, C.R., Johnson, J.A., Wilkinson, E.P., Wang, B., West, J.B., Citardi, M.J., Manwaring, K.H., Khadem, R., 2002. Implementation, calibration and accuracy testing of an image-enhanced endoscopy system. *IEEE Trans. Med. Imaging* 21 (12), 1524–1535.

- Shekhar, R., Dandekar, O., Bhat, V., Philip, M., Lei, P., Godinez, C., Sutton, E., George, I., Kavic, S., Mezrich et al., R., 2010. Live augmented reality: a new visualization method for laparoscopic surgery using continuous volumetric computed tomography. *Surg. Endosc.* 24 (8), 1976–1985.
- Sielhorst, T., Feuerstein, M., Navab, N., 2008. Advanced medical displays : a literature review of augmented reality. *J. Disp. Technol.* 4 (4), 451–467.
- Su, L.-M., Vagvolgyi, B.P., Agarwal, R., Reiley, C.E., Taylor, R.H., Hager, G.D., 2009. Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3D-CT to stereoscopic video registration. *Urology* 73 (4), 896–900.
- Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (11), 1330–1334.