



# Automatic, global registration in laparoscopic liver surgery

Bongjin Koo<sup>1</sup> · Maria R. Robu<sup>1</sup> · Moustafa Allam<sup>2</sup> · Micha Pfeiffer<sup>3</sup> · Stephen Thompson<sup>1</sup> · Kurinchi Gurusamy<sup>2</sup> · Brian Davidson<sup>2</sup> · Stefanie Speidel<sup>3</sup> · David Hawkes<sup>1</sup> · Danail Stoyanov<sup>1</sup> · Matthew J. Clarkson<sup>1</sup>

Received: 30 March 2021 / Accepted: 4 October 2021 / Published online: 26 October 2021  
© The Author(s) 2021

## Abstract

**Purpose** The initial registration of a 3D pre-operative CT model to a 2D laparoscopic video image in augmented reality systems for liver surgery needs to be fast, intuitive to perform and with minimal interruptions to the surgical intervention. Several recent methods have focussed on using easily recognisable landmarks across modalities. However, these methods still need manual annotation or manual alignment. We propose a novel, fully automatic pipeline for 3D–2D global registration in laparoscopic liver interventions.

**Methods** Firstly, we train a fully convolutional network for the semantic detection of liver contours in laparoscopic images. Secondly, we propose a novel contour-based global registration algorithm to estimate the camera pose without any manual input during surgery. The contours used are the anterior ridge and the silhouette of the liver.

**Results** We show excellent generalisation of the semantic contour detection on test data from 8 clinical cases. In quantitative experiments, the proposed contour-based registration can successfully estimate a global alignment with as little as 30% of the liver surface, a visibility ratio which is characteristic of laparoscopic interventions. Moreover, the proposed pipeline showed very promising results in clinical data from 5 laparoscopic interventions.

**Conclusions** Our proposed automatic global registration could make augmented reality systems more intuitive and usable for surgeons and easier to translate to operating rooms. Yet, as the liver is deformed significantly during surgery, it will be very beneficial to incorporate deformation into our method for more accurate registration.

**Keywords** Deep learning · Semantic contour detection · Image guidance · Augmented reality · Laparoscopy · Automatic registration

## Introduction

Augmented reality (AR) systems in laparoscopic liver surgery could help surgeons identify internal anatomical structures more clearly, especially in complex interventions. Such guidance can potentially reduce the risk of complications for the patients through safer decisions, reduced surgery time and

less blood loss. An essential component in an AR system is the registration of the pre-operative 3D liver model and the intra-operative scene in the initial stage of the intervention.

Such registration presents several challenges since the liver undergoes significant deformation due to pneumoperitoneum, it is only partially visible and it lacks reliable features [8]. Most registration algorithms can be split into two stages. Firstly, a rough global rigid transform is estimated w.r.t. the laparoscopic scene. Secondly, local alignment methods improve the results further. Multiple automatic solutions have been proposed for local alignment, assuming a good initialisation [2, 18, 20]. For liver surgery, global alignment is usually achieved manually [21, 27, 29] or in a semi-automatic way requiring annotations from the clinician during the intervention [16, 26].

An automatic registration pipeline would remove any user induced variability. It could also be repeatedly employed in order to re-initialise the registration in case of occlusion (i.e.

---

Bongjin Koo and Maria R. Robu are Joint first authors.

✉ Bongjin Koo  
b.koo@ucl.ac.uk

<sup>1</sup> Wellcome/EPSRC Centre for Interventional and Surgical Sciences and the Centre for Medical Image Computing, UCL, London, UK

<sup>2</sup> Division of Surgery and Interventional Science, UCL, London, UK

<sup>3</sup> Translational Surgical Oncology, National Center for Tumor Diseases, Dresden, Germany

due to instruments, blood) without any additional load on the clinician. Global alignment is currently the main bottleneck in automatic long-term AR systems since subsequent local registration or tracking relies on this initial stage.

While stereo laparoscopes are used mostly in robotic systems, monocular scopes are much more commonly available. As such, the rest of the paper focusses on formulating a generally applicable approach, using a single 2D laparoscopic image.

## Related works

3D–2D liver registration methods require correspondences to be found between a 3D model and a 2D image of the patient’s anatomy. The main sources of failure in the alignment estimation stem from the deformation and the partial visibility of the liver in the laparoscopic image. To tackle these challenges, prior information can be used to constrain the optimisation. The liver boundary [2], the anterior ridge and falciform ligament [15,16,20,26] have been proposed as landmarks for constraints. Currently, these techniques need manual annotation during surgery in order to obtain the liver contours [2,15,20], matching endpoints [15,20] or manual global alignment [2]. While the annotation of the liver boundary could be automated using deep learning [13], a rigid initialisation of the camera pose is still needed, which is currently achieved manually [2]. Separating the organ boundary into the anterior ridge and silhouette can lead to automating the registration since they can be matched to the corresponding contours on the 3D model [15], if a large part of both contours is visible in the image. Several approaches have been validated in synthetic experiments for partial data assuming 70–100% of the liver boundary is visible [2,20]. While such visibility ratios are achievable in open surgery [1], having an unobstructed large view of both liver lobes in laparoscopic intervention requires the cutting of the falciform ligament [15,20]. However, laparoscopic images from interventions where the falciform ligament is present show only approximately 30–50% of the liver boundary. For such cases, manual alignment is currently the only reliable option.

An alternative with promising results in the computer vision research consists of using deep learning techniques to deal with the complexity of 3D–2D registration. Such techniques have been proposed in the medical field for clinical applications where standardised datasets can be collected easily, i.e. MRI, CT, OCT scans [3]. In image guidance, such 3D–2D registration datasets are not currently available and they are extremely difficult to build. As each surgery is slightly different in pathology, organ appearance, organ geometry, patient age, type of intervention, a large dataset with multiple examples for each task is needed. Since liver surgeries are especially challenging in terms of inter-subject

variability, small datasets will result in the trained network being incapable of generalising well to new examples. For instance, a liver segmentation network trained on approximately 2000 images across 13 interventions reported poor generalisation with cases that vary too much from the training data [13].

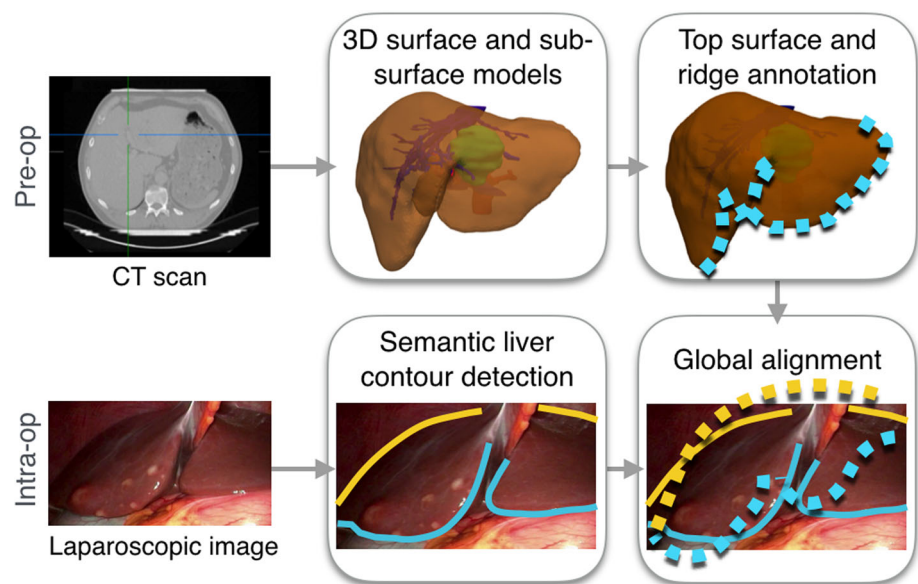
While collecting more data requires extensive manual work and data collection from multiple surgeries, a solution could be provided by training networks with synthetic data. Several approaches propose to use synthetic data for CNN-based deformation estimation from partial surfaces in 3D–3D registration [5,23]. However, building a completely simulated dataset for 3D–2D registration is extremely challenging, due to the domain gap between synthetic and real clinical data.

Recent studies propose style transfer to enhance the realism of surgical simulations [17,22]. Specifically, a synthetic dataset of photorealistic simulations of laparoscopic liver surgery is publicly available<sup>1</sup> [22]. Such large synthetic datasets are essential for advancing the current state of the art, but the issue of automatic 3D–2D registration is still not solved.

Alternatively, deep learning techniques can be used for solving well-defined tasks as part of a pipeline, such as contour detection. In the computer vision community, a real-time 3D eyelid tracking from semantic edges approach is most similar to our work [30]. They use a CNN to detect four edges of the eyelid, namely the double-fold, upper eyelid, lower eyelid and lower boundary of the bulge. These detected contours are then used to reconstruct the 3D shape and motion of the eyelids with increased realistic detail. While some similarities do exist, their method assumes the whole eye is visible at all times and explicitly uses the intersection points at the endpoints of the eyelid in the registration formulation. In laparoscopic liver surgery, such an approach would not be possible due to the partial visibility of the organ. Moreover, the variety of liver appearance and illumination fall-off due to the light source being close to the surface make the laparoscopic environment more complex. François et al. [11] propose a CNN-based framework to detect occluding contour of uterus. Occluding contours refer to boundary regions where the uterus occludes other structures and they are thus a subset of the uterus’s silhouette. This is in contrast to our method which detects the ridge as well as silhouette. This difference arises from the anatomical difference between the liver and uterus where the liver has a distinctive ridge region, but the uterus has a general spherical shape without outstanding features.

<sup>1</sup> <http://opencas.dkfz.de/image2image/>.

**Fig. 1** An overview of the proposed pre-operative and intra-operative stages for a global 3D–2D registration. Colour coding: silhouette—yellow, anterior ridge—blue



## Contributions

We propose an automatic global 3D–2D registration solution for general laparoscopic liver interventions. This work follows from a body of work utilising contours for 3D–2D liver registration [15,16,20,26], to specifically address full automation. We have developed an automated contour detection algorithm that requires no manual annotations, followed by registration. This enables fully automatic 3D–2D registration. A concurrent work [11] attempts the same goal for the surgery on the uterus.

Concretely, our contributions are as follows: Firstly, a semantic edge detection network is adapted to distinguish between different types of liver contours. Secondly, a traditional pose estimation technique is extended to match corresponding contours, which are only partially visible. We perform quantitative and qualitative experiments to assess the feasibility of the proposed method, which show promising results.

## Methods

An overview of the proposed workflow is shown in Fig. 1. The 3D surface and internal anatomical structures are segmented via a commercial service.<sup>2</sup> We also take advantage of no time limit in the pre-operative stage to pre-compute the anterior ridge and the top surface of the liver from the segmented liver mesh. These steps could be easily automated as well [24], but we chose to do it manually due to the variety in liver surface geometry when there are abnormalities present. Moreover,

the intrinsic parameters of the laparoscopic camera can also be estimated pre-operatively [32].

During surgery, there are two main components after the laparoscopic image to be registered is selected: (i) semantic liver contour detection; (ii) global 3D–2D contour-based registration. We propose to use two types of liver contours: the *anterior ridge* and *silhouette* (Fig. 1). The former is an anatomical landmark which remains fixed on the organ but can become occluded due to blood, fat or overlapping organs such as the bowel. Note that it is easy to move the liver to reveal the ridge when the bowel overlaps. The latter changes depending on the camera position and organ deformation. When used together, these contours can provide complementary constraints to the pose optimisation [15], which becomes essential under partial visibility.

### Semantic contour detection network

In the computer vision research community, the most recent approaches proposed for semantic edge detection use CNNs to achieve state-of-the-art results. We adapt CASENet [31] to predict silhouette and ridge contours of the liver, as well as background (i.e. non-liver pixels), from laparoscopic images. In addition, we pre-train the network on around 100,000 synthetic laparoscopic images because the size of our clinical dataset is very small, i.e. 133 images. This greatly helps address the overfitting on a small dataset as well as improve the generalisation capability of the network.

Once the anterior ridge and silhouette are predicted on the input laparoscopic image, they need to be matched to the corresponding contours on the pre-operative 3D liver model.

<sup>2</sup> [www.visiblepatient.com](http://www.visiblepatient.com).

### 3D–2D contour-based registration

Solutions for camera pose estimation from known 3D–2D correspondences can be obtained using well-established computer vision techniques such as Perspective-n-Point (PnP) [19]. Random Sample Consensus (RANSAC) has been proposed to deal with outliers in the correspondence set [10]. A combined PnP-RANSAC approach has been used successfully in multiple AR applications due to its computational efficiency and robustness [19].

---

#### Algorithm 1: Global contour-based 3D–2D registration

---

**Input:** Pre-op 3D liver model  $M$ ,  
 canonical liver transform  $T_C$ ,  
 2D predicted contours  $p_{label}^{2D}$ ,  
 intrinsic matrix  $K$

**Parameters:** number of initial perturbations  $m$ ,  
 maximum iterations for refinement  $i_{max}$ ,  
 maximum iterations for RANSAC  $j_{max}$ ,  
 minimum contours distance  $d_{min}$ ,  
 decay rate for refinement  $decay_{rate}$

**Result:** Global rigid transformation  $T_{global}$

```

1  $i = 0$ ;
2  $decay = 1.0$ ;
3  $T_{next} = T_C$ ;
4 for  $i < i_{max}$  do
5    $\{T_{optim}\} = \{\}$ ;
6    $\{T_{init}^C\}_m = \text{GenerateInitialGuesses}(M, T_{next}, m, decay)$ ;
7    $\{T_{init}\}_m = \text{CanonicalToOrig}(\{T_{init}^C\}_m, T_C^{-1})$ ;
8   for each  $T_{init} \in \{T_{init}\}_m$  do
9      $p_{label}^{3D} = \text{Estimate3DVisibleContours}(M, T_{init})$ ;
10     $\tilde{p}_{label}^{2D} = \text{ProjectContours2Image}(p_{label}^{3D}, T_{init}, K)$ ;
11     $\{\tilde{p}_{label}^{2D}, p_{label}^{2D}\} = \text{EstimCorresp}(\tilde{p}_{label}^{2D}, p_{label}^{2D})$ ;
12     $T_{optim} = \text{EstimPose}(M, \{\tilde{p}_{label}^{2D}, p_{label}^{2D}\}, K, j_{max}, d_{min})$ ;
13     $\{T_{optim}\} \leftarrow \text{add } T_{optim}$ ;
14   $T_{next} \leftarrow \text{save } T_{optim} \in \{T_{optim}\}$  with minimum distance error;
15   $decay = decay * decay_{rate}$ ;
16   $i = i + 1$ ;
17  $T_{global} = T_{next}$ ;

```

---

Algorithm 1 describes our proposed contour-based PnP-RANSAC extension. A transformation of the liver to a canonical space can be pre-computed ( $T_C$ ), employing the common assumption that the laparoscopic camera will be inserted through a trocar placed approximately around the belly button of the patient [15]. Let the camera follow the right-handed coordinate system with the positive  $x$ - and  $y$ -axes pointing right and down, while the positive  $z$ -axis is

pointing forwards. A range of  $m$  initial camera poses is generated in the canonical space  $\{T_{init}^C\}_m$  by random perturbations of rotation around the  $x$ -,  $y$ - and  $z$ -axes ( $\in \mathcal{N}(0, 20^\circ)$ ) of the camera (line 6), which makes the registration robust to smaller areas of the liver being visible. These initial transformation guesses  $\{T_{init}^C\}_m$  are brought back to the original space of the 3D liver model (line 7). For each initial transformation guess (line 8), the visible contours are estimated on the 3D model (line 9) for each label, i.e. anterior ridge and silhouette. Firstly, the visible surface  $M_{vis}$  is estimated from a given camera position (similar to [2]) by selecting the 3D liver model faces whose normal vector's direction is within  $\pm 90^\circ$  of the vector from the centre of the face to the camera position. Secondly, the visible surface  $M_{vis}$  is intersected with the top liver surface  $M_{top}$ . This step ensures all the points on the posterior side of the liver are excluded, since they are generally not visible in the initial stages of a laparoscopic intervention. Thirdly, the final 3D ridge points are obtained from the visible surface by  $p_{ridge}^{3D} = M_{vis} \cap M_{ridge}$  where  $M_{ridge}$  is the pre-computed ridge surface. The reason why we separate the ridge points from  $M_{vis}$  is that in some cases the top liver surface  $M_{top}$  does not fully contain  $M_{ridge}$ , i.e. the ridge line sticks out of  $M_{top}$ , depending on the annotation. Therefore, we perform  $p_{ridge}^{3D} = M_{vis} \cap M_{ridge}$  to remove the unwanted portion of the ridge line for consistency. Lastly, the silhouette  $p_{silhouette}^{3D}$  is estimated as all the boundary points of the visible surface  $M_{vis}$  that do not belong to the ridge. Once the visible 3D ridge and silhouette points are estimated, they can be projected on top of the 2D image, using the known intrinsic parameters (line 10). Correspondences between the projected and predicted contours are computed in the image space by searching for the closest neighbour and similar normals (less than  $30^\circ$  difference). The threshold used for normal similarity is to filter out correspondences where the projected and predicted contour shapes look different, even if the corresponding points are close in position, due to the deformation existing on the liver in the laparoscopic image [2]. As such, the function *EstimCorresp* on line 11 outputs a set of corresponding points  $\{\tilde{p}_{label}^{2D}, p_{label}^{2D}\}$  where the points  $\tilde{p}_{label}^{2D}$  belong to the projected 3D contours and the points  $p_{label}^{2D}$  to the predicted contours on the laparoscopic image.

The PnP-RANSAC algorithm is then employed to find the optimal camera pose  $T_{optim}$  for the current iteration (line 12). The PnP-RANSAC workflow consists of randomly selecting a minimal sample of 4 pairs from the correspondence set  $\{\tilde{p}_{label}^{2D}, p_{label}^{2D}\}$  at each iteration  $j$ . PnP is then employed to estimate the camera pose  $T_k$  for the current minimal set of point pairs. We use the P3P technique proposed in [12]. In order to measure the agreement of the whole set of correspondences with the current estimate, a distance error is computed between the projected 3D contours (transformed using  $T_k$ ) and the 2D contours. The chosen error is the modi-

fied Hausdorff distance [9] which enforces the corresponding contours to be similar. Concretely, the modified Hausdorff distance between sets  $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^{n \times 2}$  is computed as

$$\begin{aligned} \mathbf{hd}(\mathcal{X}, \mathcal{Y}) \\ = \max\{\text{mean}\{\mathbf{d}(x, \mathcal{Y}) | x \in \mathcal{X}\}, \text{mean}\{\mathbf{d}(y, \mathcal{X}) | y \in \mathcal{Y}\}\} \end{aligned} \quad (1)$$

where  $\mathbf{d}(a, \mathcal{B})$  is the minimum Euclidean distance between an element  $a \in \mathbb{R}^{n \times 2}$  and a set  $\mathcal{B} \subset \mathbb{R}^{n \times 2}$ . We compute the modified Hausdorff distance separately for the ridges and the silhouettes as we empirically found that computing the distances separately yields better registration than computing the distance together, i.e. the ridge and silhouette contours are regarded as one contour. The final distance is the sum of distances of ridge and silhouette contours. Then, the optimal camera pose is the one with the minimum distance.

On top of this PnP-RANSAC loop (lines 8–13), we introduce another loop for refinement lasting  $i_{max}$  iterations (line 4). This is to refine the estimated transformation further by starting a PnP-RANSAC loop with the optimal transformation from the previous iteration. After finishing the refinement loop, the global 3D–2D transformation  $T_{global}$  is obtained.

The U-Net-based contour extraction and PnP-RANSAC-based registration are implemented within *SmartLiver* [27, 29], a closed source application for image guided liver surgery built on top of the open-source SciKit-Surgery libraries [28].

## Experiments

### Semantic contour detection

The training clinical dataset (C) consists of 133 images extracted from two laparoscopic interventions. The source videos were recorded using NifTK's [6] IGIDataSources plugin. The data were annotated by a clinical fellow where polygonal lines were drawn on top of each contour type. The pre-training dataset (S) consists of approximately 100,000 synthetic laparoscopic images generated using [22].

Two training scenarios are considered, where the weights for the CASENet model are first initialised from ResNet50 pre-trained on the ImageNet dataset (I) [7]: (i) **I + C**: CASENet is trained on the clinical dataset (C); (ii) **I + S + C**: CASENet is pre-trained on the synthetic dataset (S), then fine-tuned on the clinical dataset (C). Data augmentation is used to make the network invariant to brightness changes, contrast, rotations, translations, scale changes and shear. When pre-training CASENet on the synthetic dataset, we use as data augmentation only brightness changes and

contrast in order to make the predictions more insensitive to different liver appearances. The Adam optimiser [14] was used for training the network with learning rate  $1 \cdot e^{-4}$ , and the training lasts for 300 epochs. A checkpoint is saved at the lowest validation loss, which is used to generate the results presented here. Train/validation set split is 80%/20%, respectively. The computation time for prediction on an image (using an NVIDIA GeForce GTX 1060 card) is around 140 milliseconds which is acceptable for use during surgery. We evaluate the performance of the proposed model on 3 test datasets: *daVinci*—9 images from a da Vinci intervention; *lap1*—9 images from 6 clinical cases; *lap2*—12 images from 1 clinical case. Figure 2 shows a selection of images from each dataset with the ground truth annotations and the predictions obtained using the two training scenarios.

Three accuracy measures are used for evaluating the network performance: *precision* P (out of all the predicted contour pixels, how many are correctly labelled?), *recall* R (how many of the ground truth contour pixels are predicted as correct?) and  $F_1$  score [4] which is defined as

$$F_1 = \frac{2 \cdot P \cdot R}{P + R + \epsilon} \quad (2)$$

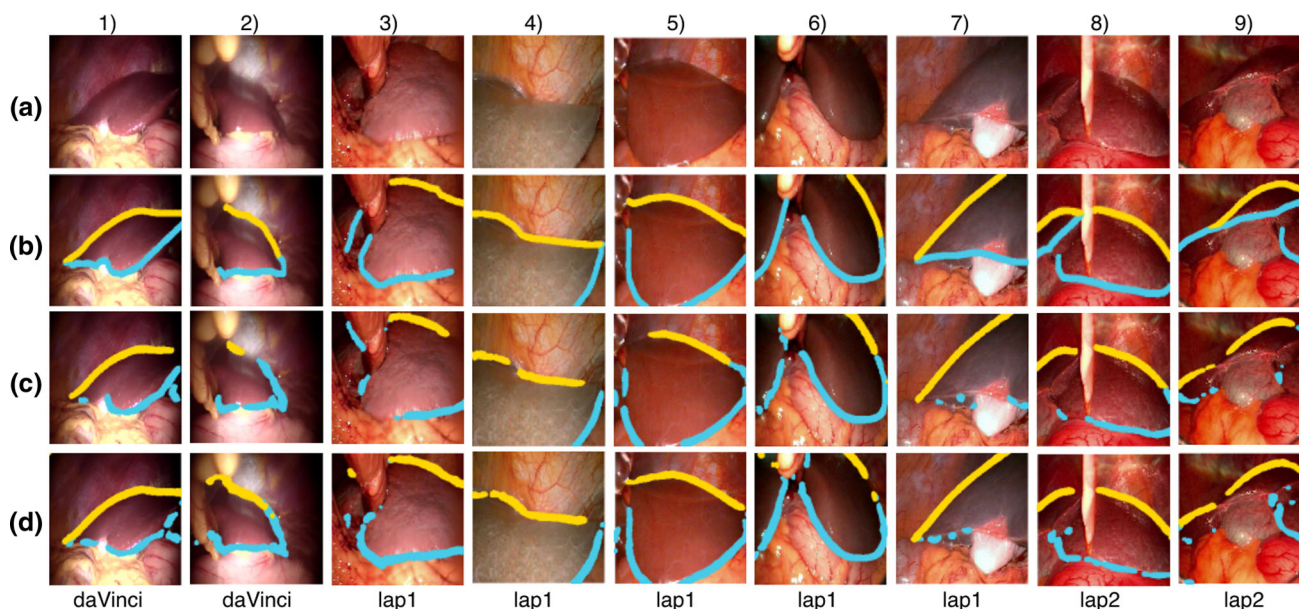
where  $\epsilon$  is a small number to avoid the denominator being zero. Table 1 summarises the results for each dataset in the two training scenarios using modified CASENet (ours) and a baseline method, U-Net. Figure 3 shows quantitative maps with true positives (green), false positives (blue) and false negatives (red) for some of the predictions in the test datasets, along with their associated  $F_1$  scores.

### Quantitative experiments

The registration performance depends on the uniqueness of the constraints imposed by the contours. Since the characteristics of the contours (such as the curvature) vary greatly depending on the viewing angle, both partial visibility and the liver region need to be taken into consideration.

We adapt the pre-operative simulation framework proposed in [25] for quantitatively analysing the performance of the proposed 3D–2D contour-based registration. Originally, the method in [25] was used for pre-operatively computing a data acquisition protocol in which the surgeon would acquire specific liver surface patches which would ultimately lead to an efficient 3D–3D registration. Their approach is appealing because it provides a way to analyse which specific camera views would result in a good registration.

The simulation framework loads a 3D liver model from a clinical case. In our synthetic experiments, 25 random camera positions are simulated on a sphere around the liver. The camera orientation is perturbed further in order for the liver not to be always at the centre of the image. For each camera, a



**Fig. 2** Qualitative results of ridge (blue) and silhouette (yellow) prediction. **a** Original images from the specified test dataset; **b** ground truth annotations; **c** predictions when CASENet was trained only on a clinical dataset (I+C); **d** predictions when CASENet was trained on synthetic and clinical dataset (I+S+C)

**Table 1** Performance of CASENet compared against a baseline, U-Net, on the test datasets measured by  $F_1$  score ( $F_1$ ), recall ( $R$ ) and precision ( $P$ )

	Dataset	Training	daVinci		lap1		lap2	
			I+C	I+S+C	I+C	I+S+C	I+C	I+S+C
U-Net	$F_1$	Ridge	0.37	0.43	0.37	0.41	0.24	0.54
		Silhouette	0.62	0.73	0.74	0.75	0.75	0.74
	$R$	Ridge	0.38	0.54	0.35	0.48	0.15	0.44
		Silhouette	0.74	0.84	0.73	0.82	0.71	0.83
	$P$	Ridge	0.39	0.38	0.43	0.37	0.72	0.70
		Silhouette	0.55	0.64	0.77	0.70	0.81	0.69
CASENet (ours)	$F_1$	Ridge	<b>0.39</b>	<b>0.46</b>	<b>0.41</b>	<b>0.45</b>	<b>0.29</b>	0.42
		Silhouette	0.61	<b>0.75</b>	0.74	<b>0.76</b>	0.73	<b>0.77</b>
	$R$	Ridge	<b>0.39</b>	0.46	<b>0.42</b>	<b>0.51</b>	<b>0.20</b>	0.32
		Silhouette	0.57	0.79	<b>0.74</b>	0.82	0.68	0.77
	$P$	Ridge	<b>0.43</b>	<b>0.48</b>	<b>0.44</b>	<b>0.45</b>	0.63	0.66
		Silhouette	<b>0.70</b>	<b>0.71</b>	0.75	<b>0.72</b>	<b>0.82</b>	<b>0.79</b>

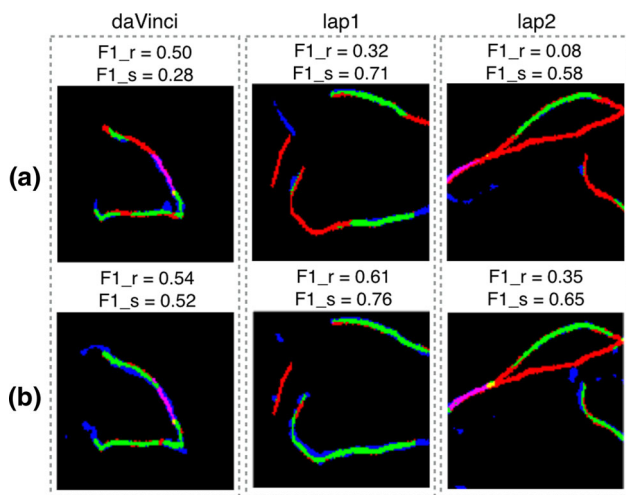
The numbers represent the average over all the images in each dataset. Higher numbers are better. (Bold numbers are when our method performs better than the baseline.) Notice that using the synthetic dataset (I+S+C) boosts the performance

synthetic image is obtained by estimating the visible contours and projecting them to 2D. For each camera position, the proposed contour-based registration is run 10 times between the synthetic image and the 3D liver surface, in order to account for the sources of randomness (lines 6 and 12 in Algorithm 1). Figure 4 shows the results. The liver visibility was computed as a ratio between the visible front liver vertices over the total vertices of the front liver surface. The root-mean-square error (RMSE) is measured between the ground truth

vertex positions of the 3D liver surface and the estimated vertices obtained after the registration process.

On top of analysing the robustness to partial visibility, such a pre-operative planning simulation can provide a clear protocol to clinicians with regard to which portions of the liver provide sufficient constraints for the registration.

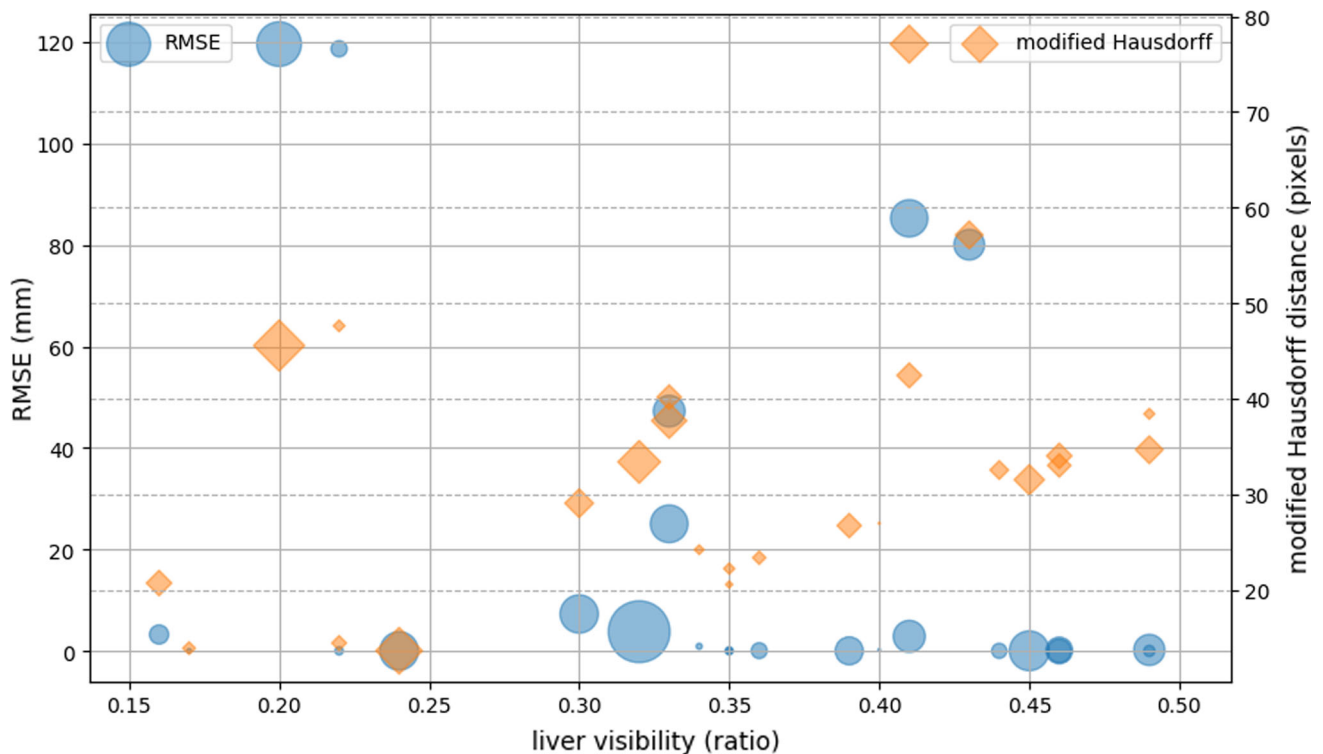
The registration takes less than 1 min which makes our method suitable for intra-operative use.



**Fig. 3** Example of  $F_1$  score maps for the ridge ( $F1_r$ ) and silhouette ( $F1_s$ ) where green = true positives, blue = false positives and red = false negatives. **a** Predictions when CASENet was trained on the clinical dataset (I + C); **b** predictions when CASENet was pre-trained on the synthetic dataset and fine-tuned on the clinical one (I+S+C)

### Qualitative experiments

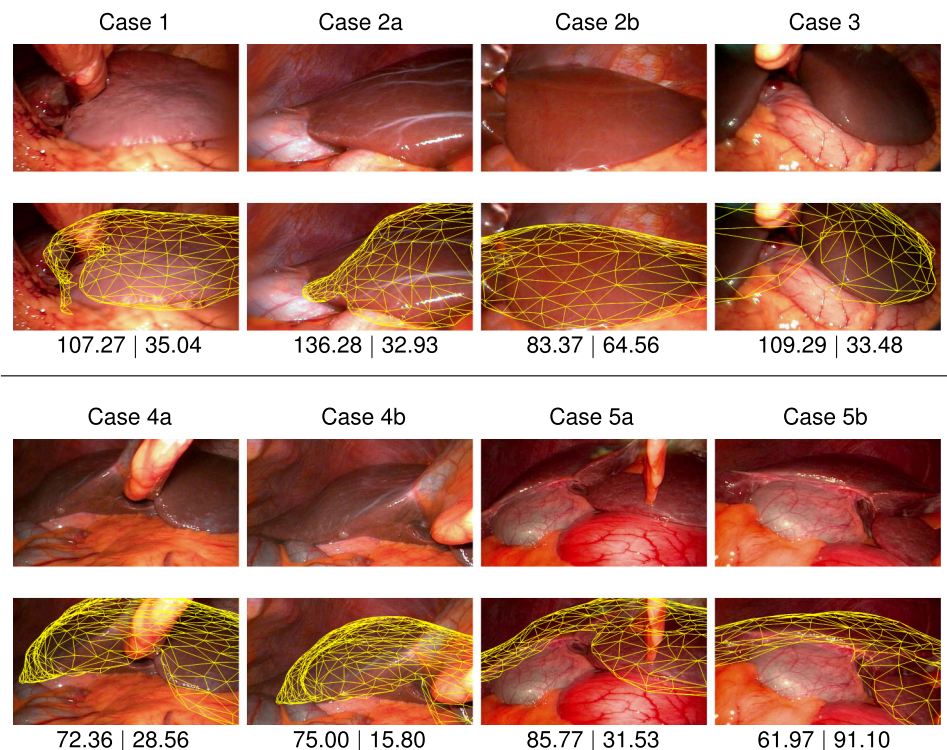
We perform experiments on real clinical data to assess the feasibility of our method in a laparoscopic liver intervention. The dataset used to validate the proposed registration pipeline is composed of 14 images from 5 retrospective clinical cases. Figure 5 illustrates the registration results where the 3D liver surface is overlaid on the input image. Without ground truth datasets for registration, we provide errors computed on the contours as well as on the vertices of the liver model against manually registered liver model. For the contours, the modified Hausdorff distance between the ground truth contours and the projected contours of the 3D liver model is computed. For RMSE, we manually register the liver model on each image and compute RMSE between all the vertices of the manually registered liver and those of the liver registered by our method. These results show the potential of our proposed registration pipeline on challenging laparoscopic images.



**Fig. 4** The registration errors (RMSEs) on the synthetic dataset along with the modified Hausdorff distances against visibility of the liver. The modified Hausdorff distance is what our method minimises for. For each trial, a synthetic image is generated from a random camera pose. Then, a 3D liver surface is registered to the image 10 times to compute the median and standard deviation (std) of the modified Hausdorff distance and RMSE. The area of each circle/diamond is proportional to the std.

(Largest area corresponds to std of 151.30 for RMSE and 51.81 for modified Hausdorff distance.) The initial RMSEs, computed at the beginning of each registration, throughout the experiments are around 250 mm. Note that the failed registrations with high RMSEs (> 45) have less than 30% (0.30 in the figure) liver, ridge or silhouette visibility (ridge and silhouette visibilities not shown in the figure)

**Fig. 5** Results of the proposed global registration pipeline on 5 retrospective clinical cases. The first row for each case shows the input laparoscopic image and second row the registered 3D liver model overlaid on the image. The numbers on the bottom row are the reprojection error in pixels (on the left) and RMSE in millimetre (on the right). The reprojection error is computed by the modified Hausdorff distance between the ground truth contours and the projected contours of the 3D liver model. RMSE is computed against the manually registered liver model's vertices



## Discussion

### Semantic contour detection

Table 1 shows how the use of synthetic dataset improves contour prediction (I+S+C). Compared to U-Net (baseline), CASENet (ours) performs better on our task and datasets. However, it is worth noting that the choice of the network architecture might not be the most critical factor for the better performance and other networks such as U-Net may suffice.

Figure 2 shows an excellent generalisation across livers with significant changes in appearance (i.e. columns 4, 6, 7) and across different image acquisition methods. Notice how pre-training improves how much of the contour gets detected, especially on the last column which presents a case never encountered in the training set C.

### Quantitative experiments

Figure 4 shows that the proposed registration method can cope with severe occlusion of the liver surface, thus occlusion of ridge and silhouette. It manages to estimate a good initial alignment (within several cm [18]) with as little as 30% visible front liver surface. Since laparoscopic images generally capture approximately 30–50% of the front liver surface, these results are highly encouraging. Notice that the failed registrations with high RMSEs ( $> 45$ ) have less than 30% (0.3 in the figure) visibility in either liver, ridge or silhouette. (Ridge and silhouette visibilities are not shown.)

### Qualitative experiments

The proposed pipeline was successful in estimating a global alignment on all 5 clinical cases in the registration dataset. Figure 5 shows the registration results where the 3D liver surface is overlaid on each input image. As observed in the figure, the proposed pipeline achieves acceptable registration for the initial registration purpose on challenging laparoscopic images with various liver geometries, appearances and viewpoints. Still, it can be observed that the deformable registration will be highly beneficial to achieve more accurate registration as the intra-operative liver shape is significantly deformed from the pre-operative one.

### Conclusion

We propose a novel fully automatic pipeline to globally register a pre-operative 3D model to a single laparoscopic image during liver interventions. The first stage involves a semantic liver contour detection network which estimates the location of the anterior ridge and the silhouette. These contours are then matched with the ones on the pre-operative 3D model in order to estimate a global rigid registration.

Validations of the proposed pipeline were conducted on synthetic and clinical data. With the synthetic data, we show that the proposed registration can estimate a global alignment with as little as 30% of the liver surface visible by extending a patient-specific pre-operative analysis. Also, the proposed



registration pipeline was successfully applied in 5 retrospective clinical cases and it was robust to the occasional errors in the contour prediction stage.

We hope the proposed automatic global registration pipeline can improve augmented reality systems in laparoscopic interventions to be more efficient and intuitive for surgeons.

**Funding** This work is supported by the CRUK (<http://www.cancerresearchuk.org>) and EPSRC Comprehensive Cancer Imaging Centre at King's College London and University College London and EPSRC CDT in Medical Imaging [EP/L016478/1]. DH, MJC, DS and ST were supported by the Wellcome/EPSRC [203145Z/16/Z]. DS receives funding from EPSRC [EP/P012841/1]. BD was supported by the NIHR Biomedical Research Centre at University College London Hospitals NHS Foundations Trust and University College London. The imaging data used for this work were obtained with funding from the Health Innovation Challenge Fund [HICF-T4-317], a parallel funding partnership between the Wellcome Trust and the Department of Health.

## Declarations

**Conflict of interest** The authors declare no conflicts of interest.

**Human and animal rights** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee.

**Consent to participate** Informed consent was obtained from all individual participants included in the study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adagolodjo Y, Golse N, Vibert E, Mathelin MD, Cotin S, Courtecuisse H (2018) Marker-based registration for large deformations—application to open liver surgery. In: ICRA, pp 1–7. <https://doi.org/10.1109/ICRA.2018.8462909>
- Adagolodjo Y, Trivisonne R, Haouchine N, Cotin S, Courtecuisse H (2017) Silhouette-based pose estimation for deformable organs application to surgical augmented reality. In: International conference on intelligent robots and systems, pp 1–7
- Altaf F, Islam SMS, Akhtar N, Janjua NK (2019) Going deep in medical image analysis: concepts, methods, challenges, and future directions. *IEEE Access* 7:99540–99572. <https://doi.org/10.1109/ACCESS.2019.2929365>
- Blair DC (1979) Information retrieval. *J Am Soc Inf Sci* 30(6):374–375. <https://doi.org/10.1002/asi.4630300621>
- Brunet JN, Mendizabal A, Petit A, Golse N, Brunet Jn, Mendizabal A, Petit A, Golse N (2019) Physics-based deep neural network for augmented reality during liver surgery. In: MICCAI. Springer International Publishing, pp 137–145
- Clarkson MJ, Zombori G, Thompson S, Totz J, Song Y, Espak M, Johnsen S, Hawkes D, Ourselin S (2015) The NifTK software platform for image-guided interventions: platform overview and NiftyLink messaging. *Int J Comput Assist Radiol Surg* 10(3):301–316. <https://doi.org/10.1007/s11548-014-1124-7>
- Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: CVPR
- dos Santos TR, Seitel A, Kilgus T, Suwelack S, Wekerle AL, Kennigott H, Speidel S, Schlemmer HP, Meinzer HP, Heimann T, Maier-Hein L (2014) Pose-independent surface matching for intra-operative soft-tissue marker-less registration. *Med Image Anal* 18(7):1101–1114
- Dubuisson MP, Jain AK (1994) A modified Hausdorff distance for object matching. In: ICPR, vol 1, pp 566–568
- Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 24(6):381–395
- François T, Calvet L, Madad ZS, Saboul D, Gasparini S, Samarakoon P, Bourdel N, Bartoli A (2020) Detecting the occluding contours of the uterus to automatise augmented laparoscopy: score, loss, dataset, evaluation and user study. *Int J Comput Assist Radiol Surg* 15:1177–1186
- Gao XS, Hou XR, Tang J, Cheng HF (2003) Complete solution classification for the perspective-three-point problem. *IEEE Trans Pattern Anal Mach Intell* 25(8):930–943
- Gibson ED, Robu MR, Thompson SA, Edwards PE, Schneider C, Gurusamy K, Davidson BR, Hawkes DJ, Barrat DC, Clarkson MJ (2017) Deep residual networks for automatic segmentation of laparoscopic videos of the liver. In: SPIE, vol 10135, pp 101351M–101351M–6
- Kingma DP, Ba J (2017) Adam: a method for stochastic optimization. *arXiv*
- Koo B, Özgür E, Le Roy B, Buc E, Bartoli A (2017) Deformable registration of a preoperative 3D liver volume to a laparoscopy image using contour and shading cues. In: MICCAI. [https://doi.org/10.1007/978-3-319-66182-7\\_38](https://doi.org/10.1007/978-3-319-66182-7_38)
- Le Roy B, Abdallah M, Espinel Y, Calvet L, Pereira B, Ozgur E, Pezet D, Buc E, Bartoli A (2020) A case series study of augmented reality in laparoscopic liver resection with a deformable preoperative model. *Surg Endosc* 34:5642–5648
- Luengo I, Flouty E, Giataganas P, Wisanuvej P, Nehme J, Stoyanov D (2018) SurReal: enhancing surgical simulation realism using style transfer. In: BMVC, p 116
- Maier-Hein L, Franz AM, dos Santos TR, Schmidt M, Fangerau M, Meinzer HP, Fitzpatrick JM (2012) Convergent iterative closest-point algorithm to accommodate anisotropic and inhomogeneous localization error. *IEEE Trans Pattern Anal Mach Intell* 34(8):1520–1532
- Marchand E, Uchiyama H, Spindler F (2016) Pose estimation for augmented reality?: A hands-on survey. *IEEE Trans Vis Comput Graph* 22:2633–2651
- Özgür E, Koo B, Le Roy B, Buc E, Bartoli A (2018) Preoperative liver registration for augmented monocular laparoscopy using backward–forward biomechanical simulation. *IJCARS* 13(10):1629–1640. <https://doi.org/10.1007/s11548-018-1842-3>
- Pessaux P, Diana M, Soler L, Piardi T, Mutter D, Marescaux J (2015) Towards cybernetic surgery: robotic and augmented reality-assisted liver segmentectomy. *Langenbeck's Arch Surg* 400(3):381–385. <https://doi.org/10.1007/s00423-014-1256-9>

22. Pfeiffer M, Funke I, Robu MR, Bodenstedt S, Engelhardt S, Roß T, Clarkson MJ, Davidson BR, Maier-Hein L, Riediger C (2019) Generating large labeled data sets for laparoscopic image processing tasks using unpaired image-to-image translation. In: MICCAI, pp 1–8
23. Pfeiffer M, Riediger C, Weitz J, Speidel S (2019) Learning soft tissue behavior of organs for surgical navigation with convolutional neural networks. *IJCARS* 14(7):1147–1155. <https://doi.org/10.1007/s11548-019-01965-7>
24. Plantefève R, Peterlik I, Haouchine N, Cotin S (2016) Patient-specific biomechanical modeling for guidance during minimally-invasive hepatic surgery. *Ann Biomed Eng.* <https://doi.org/10.1007/s10439-015-1419-z>
25. Robu MR, Edwards P, Ramalhinho J, Thompson S, Davidson B, Hawkes D, Stoyanov D, Clarkson MJ (2017) Intelligent viewpoint selection for efficient CT to video registration in laparoscopic liver surgery. *IJCARS* 12(7):1079–1088. <https://doi.org/10.1007/s11548-017-1584-7>
26. Roy BL, Ozgur E, Koo B, Buc E, Bartoli A (2019) Augmented reality guidance in laparoscopic hepatectomy with deformable semi-automatic computed tomography alignment (with video). *J Visc Surg* 156(3):261–262. <https://doi.org/10.1016/j.jviscsurg.2019.01.009>
27. Schneider C, Thompson S, Totz J, Song Y, Allam M, Sodergren M, Desjardins A, Barratt D, Ourselin S, Gurusamy K, Stoyanov D, Clarkson M, Hawkes D, Davidson B (2020) Comparison of manual and semi-automatic registration in augmented reality image-guided liver surgery: a clinical feasibility study. *Surg Endosc* 34(10):4702–4711. <https://doi.org/10.1007/s00464-020-07807-x>
28. Thompson S, Dowrick T, Ahmad M, Xiao G, Koo B, Bonmati E, Kahl K, Clarkson MJ (2020) SciKit-surgery: compact libraries for surgical navigation. *Int J Comput Assist Radiol Surg* 15(7):1075–1084. <https://doi.org/10.1007/s11548-020-02180-5>
29. Thompson S, Schneider C, Bosi M, Gurusamy K, Ourselin S, Davidson B, Hawkes D, Clarkson M (2018) In vivo estimation of target registration errors during augmented reality laparoscopic surgery. *IJCARS* 13(6):865–874. <https://doi.org/10.1007/s11548-018-1761-3>
30. Wen Q, Xu F, Lu M, Jun-Hai Y (2017) Real-time 3D eyelids tracking from semantic edges. In: ACM SIGGRAPH Asia. <https://doi.org/10.1145/3130800.3130837>
31. Yu Z, Feng C, Liu M, Ramalingam S (2017) CASENet: deep category-aware semantic edge detection. In: IEEE CVPR. <https://doi.org/10.1109/CVPR.2017.191>
32. Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22(11):1330–1334. <https://doi.org/10.1109/34.888718>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.