Automatic Preoperative 3D Model Registration in Laparoscopic Liver Resection

M. Labrunie[1,3*], M. Ribeiro[1], F. Mourthadhoi[2], C. Tilmant[1], B. Le Roy[2,3], E. Buc[1] and A. Bartoli[1,3]

[1]Universit Clermont Auvergne, Clermont Auvergne INP, CHU Clermont-Ferrand, CNRS, Institut Pascal, F-63000, Clermont-Ferrand, France.

[2]CHU Saint-Etienne, Saint-Etienne, France.

[3]SurgAR, 22 Alle Alan Turing, Clermont-Ferrand, France.

*Corresponding author(s). E-mail(s): mathieu.labrunie@etu.uca.fr;

Abstract

Purpose. Augmented Reality (AR) in Laparoscopic Liver Resection (LLR) requires anatomical landmarks and the silhouette to be found on the laparoscopic image. They are used to register the preoperative 3D model obtained from CT segmentation. The existing AR systems rely on the surgeon to 1) annotate the landmarks and silhouette and 2) provide an initial registration. These non-trivial tasks require surgeon attention which may perturb the procedure. We propose methods to solve both tasks, hence registration, automatically. **Methods.** The landmarks are the lower ridge and the falciform ligament. We solve 1) by training a U-Net from a new dataset of 1415 labelled images extracted from 68 procedures. We solve 2) by a novel automatic coarse-to-fine pose estimation method, including visibility-reasoning within an iterative robust process. In addition, we propose to divide the ridge into six anatomical sub-parts, making its annotation and use in registration more accurate. **Results.** Our method detects the silhouette with an error equivalent to an experienced surgeon. It detects the ridge and ligament with higher errors owing to under-detection. Nonetheless, our method successfully initialises the registration with tumour target registration errors of 22.4, 14.8 and 7.2 mm for 3 clinical procedures. In comparison, the errors from manual initialisation are 30.5, 15.1 and 16.3 mm. **Conclusion.** Our results are promising, suggesting that we have found an appropriate methodological approach.

Keywords: Laparoscopic Liver Resection, 3D-2D Registration, Model Initialisation, Curvilinear Landmark Detection, Monocular Vision

1 Introduction

In Laparoscopic Liver Resection (LLR), the surgeon must precisely locate the tumours in order to remove them and avoid recurrence. Because some tumours are not visible from outside the liver, additional information from preoperative CT is needed. However, combining CT and laparoscopy is challenging, especially because the liver deforms between the preoperative and intraoperative times. Augmented Reality (AR) assists the surgeon by *registering* a 3D preoperative model reconstructed from the CT to the laparoscopic image, to predict the intraoperative tumour shape and location.

We study the *registration problem* with the pipeline in figure 1. The inputs are the preoperative 3D models of the liver and tumours and the camera intrinsics are obtained by static calibration when surgery starts. Registration crucially depends on corresponding anatomical landmarks between the preoperative 3D liver model and the laparoscopic image, as well as the silhouette. The model landmarks are determined prior to surgery and the corresponding image landmarks during surgery. Existing AR systems require the surgeon to annotate the image landmarks and silhouette and to manually initialise the registration by providing a rigid pose of the 3D liver model. Both steps must be done during the procedure, which is highly inappropriate.

We present methods to overcome these issues and automate registration. Alternatively, these methods may also be used to initialise existing interactive methods. We first study the automatic detection of the image landmarks and silhouette, using a U-Net trained with a loss function adapted to thin regions. Our contribution is a database and a study of the anatomical structure of the landmarks. We define the *ridge* landmark as the lower anterior ridge of the liver and the *ligament* landmark as the junction between the falciform ligament and the liver. They are represented as curves. We fix an issue from previous works where the ridge is considered as a single curve. We propose a subdivision which copes with the variability in camera viewpoint and then a new method for automatically estimating the initial registration pose. Estimating pose automatically from curve correspondences for a single RGB image is a complex task owing to the issue of visibility, which has not received a solution in computer vision. We propose an iterative method, establishing point-wise correspondences between the landmarks, while refining visibility prediction and eventually combining the visible landmarks and the silhouette. Our methods can be readily extended to stereo images. Our experiments show that our automatic methods compare favourably with the manual results, suggesting that they could facilitate fully automatic AR in LLR.

2 Related Work

Most registration approaches split the problem into two parts: initial registration and tracking [1]. We focus on stages of the initial registration, shown in figure 1, though our methods could also be used for tracking.

Liver registration landmarks and silhouette. Registration requires correspondences between the preoperative liver and the surgical image. These are provided by the ridge and ligament landmarks (figures 2 and 3) [2]. For stereo laparoscopy, the visible surface may also be used, whereas for monocular laparoscopy, the silhouette is

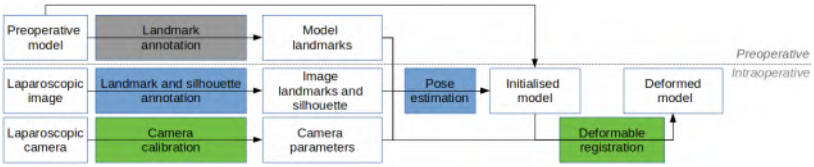


Fig. 1: Registration pipeline. The green, blue and grey boxes respectively encode the already automatic, automated by our methods and still manual tasks.

considered instead [3, 4]. The ridge is modelled as a single curve [2, 3, 5]. This does not match its real shape, as we show in a careful anatomical study in section 3.1.

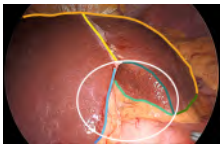
Detection of landmarks and silhouette in laparoscopic images. Detecting the landmarks and silhouette in an image is a semantic segmentation task, for which the state-of-the-art is end-to-end CNNs [6]. However, most methods deal with thick regions, as opposed to the thin curvilinear landmark shapes. Nonetheless, the U-Net architecture [6] was used on small and medium-sized datasets [7, 8] using specific function losses for thin contour detection. A compound loss with a weighted sum of Cross-Entropy (CE) and of the Dice Loss was used [8]. The Tversky Loss is a generalisation of the Dice Loss [9], more appropriate to highly imbalanced classes, which weights False Negatives (FN) more heavily than False Positives (FP). The objective is indeed not to miss the thin or small regions, the predicted labels being preferred a bit thicker than unseen. The Hybrid and Unified Focal Losses [10] are other alternatives for dealing with highly imbalanced classes. Three specific loss functions were used in successive steps [7]: CE, then a binarising loss and eventually a thinning one. We extensively tested loss combinations and report the best results in section 4.3.

Liver registration initialisation. The preoperative liver model pose is initialised manually in the vast majority of works [3, 5], with two exceptions [11, 12]. Rigid registration is performed in [11] using the ridge and a reconstructed intraoperative 3D point cloud from stereo by means of a shape matching technique. Alternatively, other 3D registration methods such as [13] could be used. Rigid registration is performed in [12] using the ridge and the silhouette. The approach is brute-force: it samples a large set of camera poses from a camera keyhole assumption and estimates the visible liver surface. In contrast, we search for the pose directly from the landmarks and silhouette, without requiring stereo, without assuming a known keyhole position and without sampling camera poses, thanks to a proposed specific highly robust process given in section 3.2.

3 Methodology

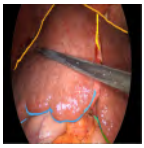
3.1 Analysis and Refinement of the Registration Landmarks

The ridge and ligament have been used as landmarks since [2]. We assume the falciform ligament was cut, which happens in most LLR procedures. Both landmarks are visible for frontal or fronto-lateral views. They are rarely entirely visible, owing

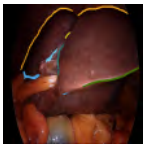


(a) Refined example, central part in the white border

(b) GT on an image from test set



(c) Predictions for the image in (b)



(d) Predictions for another test image

Fig. 2: Image, ground truth and predicted landmarks examples. Silhouette in orange, ligament in yellow, refined ridge parts in green and blue nuances.

to occlusions and self-occlusions. Previous works model the ridge as a single curve, which we call the *single ridge* model. However, depending on the camera position relative to the liver, the central part of the ridge next to the ligament may correspond to different positions in the model. We illustrate this issue in figure 2(a). Several annotations can be performed for the left central part of the ridge. This central part corresponds to the part surrounding the round ligament, close to the Rex recessus. This is the surface formed by the inferior medial segment (IVb) and the inferior lateral segment (III) outer surface closer to the intersection delimited by the frontal extremity of the ligament. This part varies substantially across patients. We propose to label both the upper and lower limits of this surface, for each side, right and left. We thus obtain the following landmarks for the central part:

- The *upper-left central limit* (may be visible from frontal left views and guessed from right fronto-lateral and frontal views)
- The *upper-right central limit* (may be visible from frontal right views and guessed from frontal left and frontal views)
- The *lower-left central limit* (may be visible from opposite lateral views)
- The *lower-right central limit* (may be visible from opposite lateral views)

We call this the *split ridge* model. Its most reliable landmarks are the *right ridge* and *left ridge* from each side of the central part. We illustrate this delimitation in figure 2.

3.2 Automatic Registration Initialisation

Iterative deformable registration requires an initialisation. This is often obtained with a rigid model, which consists in estimating the camera pose with respect to the liver model. This is typically done manually. Indeed, Perspective-n-Point (PnP) pose estimation cannot be directly used because 3D-2D point correspondences are not known between the preoperative 3D model and the 2D image landmarks.

We propose to solve this challenging pose problem iteratively, in three coarse to fine steps, as given by Algorithm 1. The main differences between these steps are the preoperative 3D vertices (with their image correspondences) selected for pose estimation in a RANSAC-based PnP solution. The coarse step ① uses all landmark vertices. The following step ② refines pose iteratively using the visible landmark vertices only, determined from the current pose estimate (line 34, figure 3(c)). The fine step ③ adds the silhouette vertices related to the image silhouette (line 32).

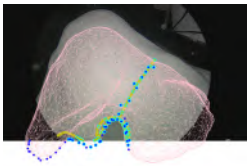
Algorithm 1 Pose Estimation Pseudo-code

```
1:  main variables:
2:  V_L: 3D landmark vertex coordinates in preoperative space
3:  I_L, I_S: 2D landmark and silhouette coordinates in image space
4:  T: threshold set for inlier selection in extended RANSAC-based PnP
5:
6:  MinError ← +∞, BestPose ← ∅
7:  ① Coarse estimation from all landmark vertices
8:  BestPose, MinError ← EstimatePose(V_L,I_L,T,MinError,BestPose)
9:  ② Refinement from visible landmark vertices only
10: BestPose, MinError ← RefinePose(False,V_L,I_L,I_S,T,MinError,BestPose)
11: ③ Refinement from visible landmark vertices and the silhouette
12: BestPose, MinError ← RefinePose(True,V_L,I_L,I_S,T,MinError,BestPose)
13:
14: procedure EstimatePose(V_t,I_t,T,MinError,InitialPose)
15:     BestPose ← InitialPose
16:     Î_t ← SampleImageLandmarksAsModelVertexOnes(I_t,V_t)
17:     for each τ in T do
18:         EstimatedPose ← RANSACPnP(V_t,Î_t,τ,InitialPose)
19:         ProjectedV_t ← ProjectToImagePlane(V_t,EstimatedPose)
20:         MSD ← ComputeMSD(GetVisibleProjections(ProjectedV_t),I_t)
21:         if MSD<MinError then MinError←MSD, BestPose ← EstimatedPose
22:     return BestPose, MinError
23:
24: procedure RefinePose(WithSilhouette,V_L,I_L,I_S,T,MinError,BestPose)
25:     PreviousMinError ← +∞
26:     while MinError<PreviousMinError do
27:         PreviousMinError ← MinError
28:         VisibleV_L ← DetermineVisibleVerticesInImage(V_L,BestPose)
29:         if WithSilhouette then
30:             V_S ← DetermineSilhouetteVertices(BestPose)
31:             V̂_S, Î_S ← GetCorrespondences(V_S,I_S,BestPose)
32:             V_t ← [VisibleV_L, V̂_S], I_t ← [I_L, Î_S]
33:         else
34:             V_t ← VisibleV_L, I_t ← I_L
35:         BestPose,MinError ← EstimatePose(V_t,I_t,T,MinError,BestPose)
36:     return BestPose, MinError
```

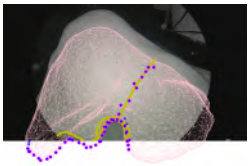
The correspondences are the closest points between the projection of the silhouette vertices and the image silhouette ($\hat{V}_S$ and $\hat{I}_S$ in line 31).

The procedure for estimating the pose is given in lines 14-22. First, the landmark image curves are sampled uniformly to create one-to-one correspondences with the model vertices (line 16). From these 3D-2D correspondences, we solve a RANSAC-based PnP (line 18) using OpenCV. PnP uses a non-linear Levenberg-Marquardt minimisation to refine the initial pose. For ①, this is initialised with Direct Linear Transformation while for ② and ③ it uses the best pose from the previous steps.

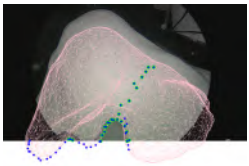
RANSAC highly depends on the reprojection error threshold defining the inlier set. An unadapted threshold leads to an inlier set with too many or too few points and then to an incorrect pose. In the problem at hand, the threshold cannot be chosen a priori. This is because the error not only depends on observation noise, as in classical vision problems, but also on modelling error. The latter stems from several factors, the use of a rigid registration to approximate the real deformation field being the strongest one. We propose an extension of RANSAC to determine an optimal threshold at runtime. This works by repeating RANSAC for several thresholds from a set $T \subset \mathbb{R}$. Each threshold $\tau \in T$ gives an inlier set (figure 3(a)) and a pose estimate. We eventually select the best solution using the Mean Sum of Distances



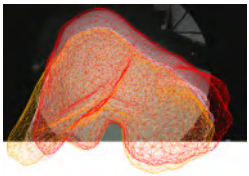
(a) Original image and model landmarks (yellow and blue) used in ① (coarse pose estimation) for solving PnP. In pink, best pose result for ① w.r.t. MSD after inlier selection (cyan) by RANSAC.



(b) Example of image and model landmarks (respectively magenta and yellow) used for MSD computation in one of the first two steps (① and ②).



(c) Example of visible landmarks (lime) used as inputs in ②.



(d) Example of refined pose estimate after step ② and final one ③ (respectively orange and red).

Fig. 3: Illustration of the proposed method details for automatic pose estimation.

(MSD) criterion [14] between the image landmarks I_t and the reprojected visible model vertices (line 20, figure 3(b)). The MSD between point sets $\boldsymbol{U} = (u_1, ..., u_n)$ and $\boldsymbol{V} = (v_1, ..., v_m)$ is:

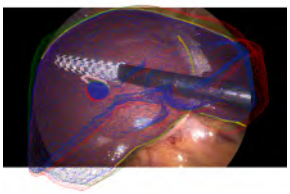
$$\text{MSD}(\boldsymbol{U}, \boldsymbol{V}) = \frac{1}{m+n} \left(\sum_{i=1}^{m} \min_j \|v_i - u_j\| + \sum_{i=1}^{n} \min_j \|u_i - v_j\| \right).$$

Usually, $n = m$ so the measure is symmetric. The mean may possibly replace the sum for each component. Unlike line 28, only self-occluded landmark vertices are considered invisible. They are determined by a ray-triangle intersection method from trimesh. Projected landmark vertices in the black border and outside the image are kept in the MSD computation. Figure 3(d) shows results for each step of our pipeline.

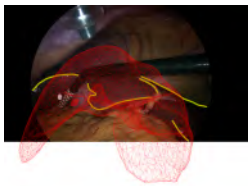
4 Experimental Evaluation

4.1 Overview

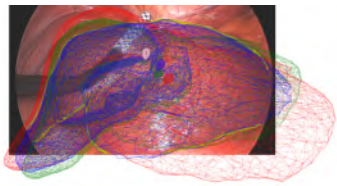
We first find a baseline objective for automatic landmark and silhouette detection by determining the manual annotation interoperator variability. We then evaluate automatic annotation detection and registration initialisation for both the single and split ridge models. We merge the left ridge and upper left central limit as well as the right ridge and right central limit to replicate the single ridge model. Finally, we evaluate the pose obtained from our automatic method against the optimal and manual poses in phantom and clinical datasets. We also compare to deformable registration initialised from these poses to achieve the whole registration pipeline. All experiments



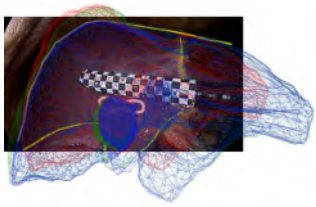
(a) Patient 1, TRE 1/2/3: 13.0/12.9/10.2



(b) Patient 2, failure case, TRE 1: 39.4



(c) Patient 3, TRE 1/2/3: 32.5/22.2/20.2



(d) Patient 4, TRE 1/2/3: 18.4/7.0/6.7

Fig. 4: Manual (1-red) and automatic (2-green) pose reprojection for the liver and tumour. Deformable registration reprojection (3-blue) initialised from 2. The tumour GT, landmarks and silhouette are in pink and yellow. The TRE is in mm.

were performed on an Intel Core i9-10900K CPU @ 3.70GHz 20 computer with graphics card GeForce RTX 2080 Ti.

4.2 Manual Annotation Interoperator Variability

We construct a dataset where 5 surgeons annotated the landmarks and silhouette for 8 laparoscopic 1080p images from different patients. For each image we compute the average closest distance across annotations, equivalent to the MSD criterion, as well as the percentage of annotations within 15 px and 30 px agreement. The estimated mean manual annotation interoperator variability is slightly over 30 px, see table 1. However, some images substantially increase variability as can be seen with the median being around 20 px. More than 80% of each landmark or silhouette from one annotation are within 30 px from another annotation. These are the baseline objectives for automatic detection.

Table 1: Manual annotation variability between 5 surgeons (2 junior and 3 senior ones) on 8 images, measured by mean, median and max MSD across images and mean percentages of annotation within 15 and 30 px for the landmarks and silhouette.

	Mean MSD (px)	Median MSD (px)	Max MSD (px)	% < 15 px	% < 30 px
Silhouette	32	19	102	83	87
Ligament	29	17	96	68	80
Single Ridge	33	20	86	70	81

4.3 Automatic Landmark and Silhouette Detection

Between 10 and 30 1080p frames were extracted from LLR videos of 68 patients. The frames were selected to represent various views and configurations of each liver. Manual annotations of the landmarks and the silhouette were performed on each image. Our dataset contains 1415 annotated images. We used Leave One (patient) Out Cross-Validation (LOOCV) to evaluate detection accuracy. For each fold, train and validation sets respectively use 90% and 10% of the randomly selected data. The test set uses the images from the left-out patient. Training takes about 20 min for a fold, while detection takes less than 10 ms. In contrast, the manual annotation of all the landmarks and silhouette can take between 30 s and about 2 min for each image.

We evaluate the detection using both components of the MSD criterion: the mean closest distance of predictions to GT and the mean closest distance of GT to predictions, whose mean is the MSD. We compute the topological skeleton of the predictions to form 1-pixel thick contours. We compute adapted sensitivity and precision, which are not well defined for thin regions. The adapted versions use dilated labels with disc radii of 15 and 30 px. The GT contours are dilated for precision computation, while the predicted ones are dilated to compute sensitivity.

We use an end-to-end U-Net [6] with input image size of 256×256. We tested the state-of-the-art loss functions for thin regions and imbalanced classes described in section 2. In short, the best results for thin outputs are obtained for the compound CE+Tversky loss, with respective weights of 5 and 1 for the CE and Tversky losses as well as 0.05 and 0.95 for the FP and FN components of the Tversky loss. Figures 2(c) and 2(d) illustrate results on test images. It performs well for the silhouette while missing parts for the ridge and the ligament.

Table 2: Detection accuracy for a U-Net trained with a CE+Tversky loss in an LOOCV manner on 68 patients combining 1415 images. Mean, median and max of MSD components and adapted sensitivity and precision are computed across patients (left-out test sets) for the landmarks and silhouette with single and split ridges.

	Distances (px) predicted from GT - Mean/ Median/Max	Distances (px) GT from predicted - Mean/ Median/Max	Sensitivity - Mean/ Median/Min (%) 15 and 30 px tolerance radius	Precision - Mean/ Median/Min (%) 15 and 30 px tolerance radius
Silhouette from split	35/22/181	51/30/332	80/83/8 85/88/12	84/89/6 87/91/11
Silhouette from single	39/23/224	56/33/383	81/85/09 86/89/13	83/88/05 87/91/11
Ligament from split	77/56/368	53/22/358	55/50/15 66/66/18	76/81/07 84/91/12
Ligament from single	58/33/238	37/15/295	58/60/10 70/72/20	76/83/05 88/92/09
Merged Ridge from split	140/102/532	97/55/485	42/41/4 51/50/7	66/65/17 80/87/23
Ridge from single	79/64/296	88/64/362	50/53/05 61/66/09	61/64/05 72/76/07

Table 2 summarises the results. For both the single and split ridge models, the silhouette brings the error close to the manual annotation interoperator variability, of 30 and 20 px for the mean and median MSD. Over 85% of the predicted points are within 30 px of GT and over 85% of the GT points are predicted within 30 px. This confirms that the method handles the silhouette very well. For the ligament, the error is slightly over. However, about 30% of the GT points are missed with a 30 px tolerance, from the single ridge model. The error is larger from the split one, suggesting that the ligament can be mistaken for ridge subparts. The ridge, whether split or not, obtains larger errors though using the single one allows fewer misdetection. The detection method thus slightly outperforms with the single ridge model. The variability is large: some patients have very large errors, which is explained by the high deformations induced by tools, creating misdetection, partial views and blurry images. Modifying the loss function and adding geometric or colour data augmentation did not improve the overall results.

4.4 Automatic Registration Initialisation

The pose estimation method is first evaluated on the phantom dataset from [15]. A preoperative model subject to 10 non-rigid synthetic deformations was 3D printed for each deformation. 10 views were captured for each, along with the camera parameters, and the optimal poses were computed. The surface GT mesh is thus known for each image. We use the Mean Absolute distance Error (MAE) between the vector sets of 3D GT surface node positions $\boldsymbol{U}$ and of the estimated ones $\hat{\boldsymbol{U}}$ as evaluation metric $\mathrm{MAE}(\boldsymbol{U}, \hat{\boldsymbol{U}}) \leftarrow \frac{1}{N}\sum_{i=1}^{N}\|u_i - \hat{u}_i\|$. The orientation error angle $\theta \leftarrow \arccos\frac{\mathrm{Tr}\,(R\hat{R}^T)-1}{2}$ is computed from the optimal and estimated pose rotation matrices $R, \hat{R} \in SO(3)$. The landmarks and silhouette are manually annotated for each image and the model. The method is then evaluated on a clinical dataset from [16]. GT tumour profiles were obtained from an ultrasound probe in 4 laparascopic procedures, each dealing with tens of images. We compute the Target Registration Error (TRE) between each GT tumour profile and the predicted tumour volume as [16]. We use the annotations and manual initialisations that were provided in this dataset. For both datasets, we use a baseline deformable registration algorithm that reproduces [17], without the point cloud term. This becomes an energy-based optimisation based on a landmark and silhouette data term and a model's internal energy term, with respective weights 0.1 and 1. We use Saint-Venant Kirchhoff deformation modelling and dimension reduction by Locally Linear Embedding [18], keeping 100 then 200 components in a stiff-to-flexible strategy. T is composed of 6 threshold values between 0.05% and 25% of the image diagonal (1-500 px in 1080p images).

Tables 3 and 4 summarise the results obtained on both datasets. For the phantom dataset, we first compute MAE between GT and both optimal and predicted poses. The predicted pose for the proposed split ridge model obtains mean and median errors of respectively 15.9 mm and 11.6 mm in comparison to the 7.7 mm and 7.4 mm ones of the optimal pose. The single ridge model performs substantially worse due to the wrong correspondences in fronto-lateral views. We can see that each refinement step

Table 3: Pose estimation accuracy on a phantom dataset. MAE statistics are estimated across images between predicted pose from a) split and b) single ridge models and GT for each algorithm step. Optimal pose and registration initialised from both poses are also evaluated. The rotation angle between predicted and optimal poses is computed.

	GT vs optimal pose	GT vs predicted pose a) split ridge model b) single ridge model	Optimal vs predicted pose a)	Registration from optimal pose vs GT	Registration from automatic initialisation a) vs GT
Error	MAE (mm)	MAE (mm) Steps ①/②/③	θ () ①/②/③	MAE (mm)	MAE (mm)
Mean	$7.7^{\pm 0.7}$	$a)68.8/36.8/15.9^{\pm 2.9}$ $b)178.5/109.8/95^{\pm 28.6}$	29/20/11	$8.1^{\pm 0.7}$	$13.1^{\pm 2.8}$
Median	7.4	$a)51.0/24.0/11.6$ $b)108.3/66.9/19.7$	23/14/8	7.3	10.0
$\eta_{.90}$	13.5	$a)112.2/69.5/27.1$ $b)359.5/291.1/338.1$	45/39/17	13.3	19.8

improves the previous result. This is also the case for the orientation which gets progressively closer to the optimal one. The largest errors are obtained for side views where very few landmarks are visible, and should not be considered in actual liver laparoscopy. This demonstrates the relevance of our proposed automatic initialisation method, only based on single image landmarks and silhouette. Using the baseline registration method, the final errors starting from the optimal and predicted poses have about 5 mm mean difference. This last comparison should be considered cautiously as it depends on the registration method and parameters. However, 90% of the views have an MAE lower than 2 cm after registration, which is promising.

Pose estimation performance is confirmed in the clinical dataset [16], where our automatic pose leads to a tumour TRE lower than or equal to the manual initialisation and always lower than 2.3 cm. Deformable registration from these initialisations further reduces the error. Figure 4 illustrates the results for all patients. Automatic initialisation only fails for Patient 2 (e.g. figure 4(b))), owing to a strong non-rigid deformation induced by the ultrasound probe and the narrowness of the views in this procedure, making the PnP problem ill-posed. The fully automatic initialisation process takes between 10 and 40 seconds.

5 Conclusion

We have proposed methods to achieve fully automatic registration of a preoperative 3D liver model to a laparoscopic image. We bring the new largest manually annotated surgical liver dataset, an analysis of the landmarks, a landmark and silhouette detector and an extended robust pose estimator capable to estimate its own inlier cut-off threshold. Our experimental results show that these methods work well on clinical cases, reducing tumour TRE compared to the current manual methods. This resolves a major limitation of current methods requiring surgeon attention, which is not feasible in clinical practice. We have planned to improve automatic curvilinear landmark detection by expanding our dataset, using semi-synthetic data, exploring thicker landmark models and subdividing the ridge after detection. This will allow the evaluation

Table 4: Pose estimation accuracy on a clinical dataset. TRE is evaluated for both estimated initialisations, i.e. manual and automatic, and their following registrations.

Error for Patients 1/3/4 (mm)	Between manual pose and GT	Between automatic estimated pose and GT	Between registration result from manual pose and GT	Between registration result from automatic estimated pose and GT
TRE	15.1/30.5/16.3	14.8/22.4/7.2	13.4/28.9/16.0	12.5/20.8/6.9

and clinical validation of the full pipeline to be conducted, including the comparison of several baseline deformable registration methods. We will extend the pipeline to multiple image registration and use our methods to tackle the next fundamental step of real time liver tracking.

Declarations

M. Labrunie is supported by a CIFRE PhD fellowship (N2021/0184) from ANRT under partnership between EnCoV and SurgAR. The authors declare no conflicts of interest. Our study received ethical approval (IRB00008526-2019-CE58) issued by CPP Sud-Est VI in Clermont-Ferrand, France. Informed consent was obtained from all individual participants included in the study.

References

[1] Bernhardt, S., Nicolau, S.A., Soler, L., Doignon, C.: The status of augmented reality in laparoscopic surgery as of 2016. Med Image Anal **37**, 66–90 (2017)

[2] Plantefève, R., Haouchine, N., Radoux, J.-P., Cotin, S.: Automatic alignment of pre and intraoperative data using anatomical landmarks for augmented laparoscopic liver surgery. In: Biomedical Simulation (2014)

[3] Özgür, E., Koo, B., Le Roy, B., Buc, E., Bartoli, A.: Preoperative liver registration for augmented monocular laparoscopy using backward-forward biomechanical simulation. IJCARS **13**(10), 1629–1640 (2018)

[4] Adagolodjo, Y., Trivisonne, R., Haouchine, N., Cotin, S., Courtecuisse, H.: Silhouette-based pose estimation for deformable organs application to surgical augmented reality. In: IROS (2017)

[5] Espinel, Y., Özgür, E., Calvet, L., Le Roy, B., Buc, E., Bartoli, A.: Combining Visual Cues with Interactions for 3D-2D Registration in Liver Laparoscopy. Ann Biomed Eng **48**(6), 1712–1727 (2020)

[6] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)

[7] Franois, T., Calvet, L., Madad Zadeh, S., Saboul, D., Gasparini, S., Samarakoon, P., Bourdel, N., Bartoli, A.: Detecting the occluding contours of the

uterus to automatise augmented laparoscopy: score, loss, dataset, evaluation and user study. IJCARS **15**(7), 1177–1186 (2020)

[8] Zhu, J., Styler, W., Calloway, I.: A CNN-based tool for automatic tongue contour tracking in ultrasound images. Preprint at https://arxiv.org/abs/1907.10210 (2019)

[9] Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: Machine Learning in Medical Imaging (2017)

[10] Yeung, M., Sala, E., Schnlieb, C.-B., Rundo, L.: Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Preprint at https://arxiv.org/abs/2102.04525 (2021)

[11] Robu, M.R., Ramalhinho, J., Thompson, S., Gurusamy, K., Davidson, B., Hawkes, D., Stoyanov, D., Clarkson, M.J.: Global rigid registration of CT to video in laparoscopic liver surgery. IJCARS **13**(6), 947–956 (2018)

[12] Koo, B., Robu, M.R., Allam, M., Pfeiffer, M., Thompson, S., Gurusamy, K., Davidson, B., Speidel, S., Hawkes, D., Stoyanov, D., Clarkson, M.J.: Automatic, global registration in laparoscopic liver surgery. IJCARS (2021)

[13] Min, Z., Liu, L., Meng, M.Q.-H.: Generalized non-rigid point set registration with hybrid mixture models considering anisotropic positional uncertainties. In: MICCAI (2019)

[14] Li, M., Kambhamettu, C., Stone, M.: Automatic contour tracking in ultrasound images. Clin Linguist Phon **19**(6-7), 545–554 (2005)

[15] Espinel, Y., Calvet, L., Botros, K., Buc, E., Tilmant, C., Bartoli, A.: Using multiple images and contours for deformable 3D-2D registration of a preoperative ct in laparoscopic liver surgery. In: MICCAI (2021)

[16] Rabbani, N., Calvet, L., Espinel, Y., Le Roy, B., Ribeiro, M., Buc, E., Bartoli, A.: A methodology and clinical dataset with ground-truth to evaluate registration accuracy quantitatively in computer-assisted laparoscopic liver resection. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 1–10 (2021)

[17] Collins, T., Pizarro, D., Gasparini, S., Bourdel, N., Chauvet, P., Canis, M., Calvet, L., Bartoli, A.: Augmented Reality Guided Laparoscopic Surgery of the Uterus. IEEE Trans Med Imaging **40**(1), 371–380 (2021)

[18] Modrzejewski, R.: Recalage déformable, jeux de données et protocoles d'évaluation pour la chirurgie mini-invasive abdominale augmentée. PhD thesis, Université Clermont Auvergne (August 2020)