Automatic 3D/2D Deformable Registration in Minimally Invasive Liver Resection using a Mesh Recovery Network

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

We propose the patient-specific Liver Mesh Recovery (LMR) framework, to automatically achieve Augmented Reality (AR) guidance by registering a preoperative 3D model in Minimally Invasive Liver Resection (MILR). Existing methods solve registration in MILR by pose estimation followed with numerical optimisation and suffer from a prohibitive intraoperative runtime. The proposed LMR is inspired by the recent Human Mesh Recovery (HMR) framework and forms the first learning-based method to solve registration in MILR. In contrast to existing methods, the computation load in LMR occurs preoperatively, at training time. We construct a patient-specific deformation model and generate patient-specific training data reproducing the typical defects of the automatically detected registration primitives. Experimental results show that LMR's registration accuracy is on par with optimisation-based methods, whilst running in real-time intraoperatively.

Keywords: Mini-invasive Surgery, Liver, 3D/2D Registration, Augmented Reality

1. Introduction

Minimally Invasive Liver Resection (MILR) is the surgical removal of liver tumours by laparoscopy or robot-assisted intervention. It has tremendous advantages over open surgery, reducing recovery time and the risk of postoperative complications. However, the intraoperative tumour localisation remains extremely challenging, for two main reasons. First, the mini-invasive setup prevents the surgeon to palpate the liver. Second, the liver tissue is soft, causing strong deformation under pneumoperitoneum, breathing, and mobilisation. Importantly, a preoperative volume (an abdominal CT or MRI), primarily used for diagnosis, is routinely available which clearly shows the tumours. It is however not possible to mentally transfer the tumour locations to the intraoperative field with sufficient accuracy (Adballah et al., 2022). This has motivated the development of computer assistance using Augmented Reality (AR), which combines the preoperative volume with the intraoperative images in a virtual see-through manner to reveal the tumour locations. The key problem of transferring information from the preoperative volume to the intraoperative images must thus be solved by a computational method, which currently represents the main challenge to achieve AR in MILR. Specifically, the key problem is called *registration*, which establishes a non-rigid transformation from the preoperative volume to the intraoperative images, generally decomposed into rigid and non-rigid parts, respectively the pose and the deformation. Existing registration methods in MILR (Özgür et al., 2018; Koo et al., 2022; Labrunie et al., 2022) rely on landmark correspondences, namely the lower ridge and the falciform ligament (appendix figures 3 and 4). In the preoperative stage, the liver volume is segmented to reconstruct a virtual preoperative 3D model including the tumours and the landmarks. In

© 2023 Anonymous.

the intraoperative stage, the current image is processed to detect the landmarks and the liver silhouette, called the primitives. The pose and the deformation are then computed from the primitives, the latter also including biomechanical constraints. Existing methods mainly carry out the computation using nonconvex numerical optimisation, typically using a gradient-based method such as Gauss-Newton. Together, they thus form the *optimisation-based approach* to registration in MILR. The state-of-the-art accuracy measured by tumour Target Registration Error (TRE) is typically within a few centimeters, which was shown to be clinically relevant (Zhong et al., 2017). However, these methods have a main structural limitation, which is that most of the computation is done intraoperatively, leading to a prohibitive intraoperative runtime. Indeed, the intraoperative optimisation process involves $\mathcal{O}(100)$ parameters and a nonconvex cost function, typically requiring dozens of iterations to convergence. This strongly limits the usability of these methods to achieve AR in MILR.

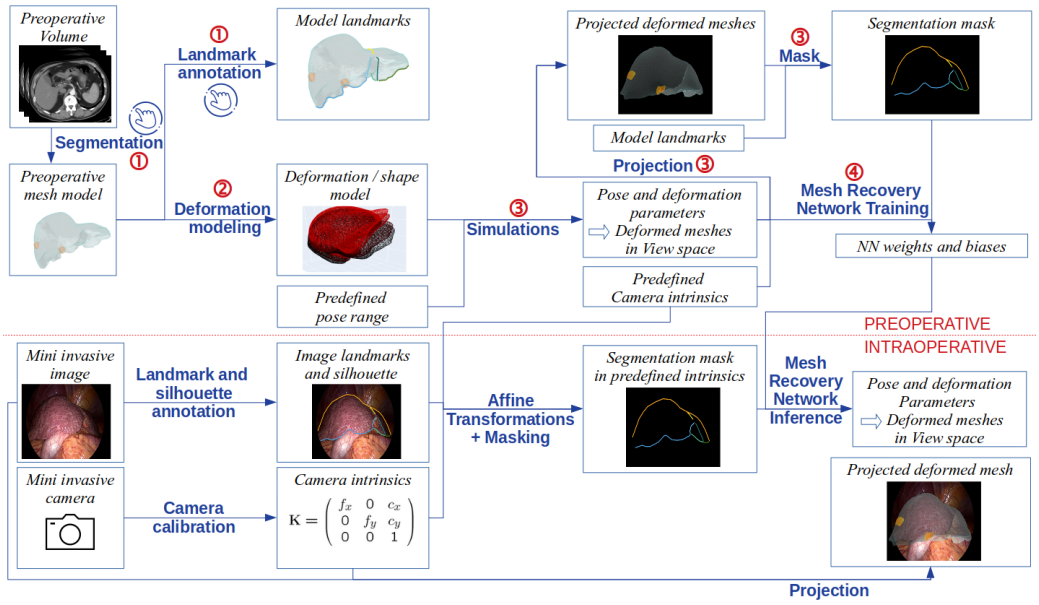


Figure 1: Proposed patient-specific registration pipeline in MILR. Intraoperatively, a mask of primitives, annotated from the current image and transformed to a predefined space, is input to the network in order to obtain the related deformed mesh. Preoperative steps provide essential data such as the preoperative mesh, a deformation model and a pose range. Deformed meshes and associated segmentation masks obtained from simulated pose and deformation enable to train the network.

We propose the Liver Mesh Recovery (LMR) framework, forming a novel *learning-based approach* to registration in MILR. LMR is inspired by recent methods to reconstruct the human body as a 3D mesh from a single image (Kanazawa et al., 2018; Tian et al., 2022). LMR directly regresses the registration from the intraoperative image primitives. Specifically, the registration is represented by pose and deformation parameters that define the intraoperative liver state. A major difficulty is the absence of labelled patient-generic datasets and

the fact that the real patient-specific images only become available at the time of surgery. Training from these images is thus not possible, first, as the registration labels cannot be reliably found and, second, as surgery has already started. A related difficulty is that the intrinsic camera parameters also become available only when surgery starts. We propose a novel key idea, which is to train preoperatively from patient-specific semi-synthetic data generated from the preoperative 3D model. We bring three main technical contributions. First, the LMR framework, including a learned iterative refinement loop. Second, a physics-based rendering method which generates patient-specific image primitive masks for a range of pose and deformation incorporating the typical defects of automatic primitive detectors. Third, a process to train LMR with default intrinsic camera parameters and the adaptation to their actual value at inference time. In contrast to the optimisation-based approach, the proposed LMR transfers most of the computational load to the preoperative stage, making intraoperative registration fast and fully automatic. We report experimental results showing that LMR's accuracy is on par with the state-of-the-art, thus clinically relevant, while running in real-time intraoperatively.

2. Related Work

2.1. Registration in Minimally Invasive Liver Resection

Existing methods use the same landmarks (Plantefeve et al., 2014) and the silhouette (Koo et al., 2017). The methods generally follow two steps, shown in appendix figure 4: 1) pose estimation, forming an initial rigid registration, and 2) deformation estimation. Pose estimation requires manual interaction (Espinel et al., 2020; Özgür et al., 2018), though recent work attempted automatic solutions. The method in (Robu et al., 2018) uses a shape-matching technique with the ridge and a 3D point cloud reconstructed from stereo. The method in (Koo et al., 2022) samples the pose space and evaluates the reprojection error in the ridge and the silhouette. The method in (Labrunie et al., 2022) searches for the pose directly from the primitives using a visibility reasoning to determine landmark correspondences. The deformation is estimated by means of intraoperative numerical optimisation of a compound cost function, with a main term exploiting the primitives. Optional terms were tried, such as shading (Koo et al., 2017). The deformation is regularised from the strain energy of a constitutive model and generic biomechanical parameters. Low-dimensional deformation models such as Locally Linear Embedding (LLE) (Roweis and Saul, 2000) were also incorporated to stabilise and speed up optimisation (Modrzejewski, 2020; Labrunie et al., 2022).

2.2. Human Body Shape Reconstruction

Human body shape reconstruction uses a single image to estimate a physically plausible mesh aligned with the image. Most methods combine a reference shape model and joint pose (Loper et al., 2015) to model the mesh. Computationally, both the optimisation-based and learning-based approaches were studied (Tian et al., 2022). Optimisation-based methods first compute an initial pose and shape, followed by an optimisation of the shape parameters for multiple image cues, including the silhouette (Guan et al., 2009). Learning-based methods are recent and stem from the end-to-end Human Mesh Recovery (HMR)

framework (Kanazawa et al., 2018). HMR starts with a CNN encoder such as ResNet-50 to extract features. It then concatenates the features with the pose, shape and camera parameters, which are all fed to a regression network with two fully-connected layers. The regression network uses Iterative Error Feedback (IEF) (Carreira et al., 2016) to update the parameters iteratively. At training, the image joint locations are known. The reprojection of the estimated 3D joints allows one to compute a 2D reprojection loss. When 3D Ground Truth (GT) is known, a 3D loss is computed directly on the 3D joints as well as on the parameters. In addition, a discriminator is used to improve the shape and pose plausibility, which is especially important in the absence of 3D GT.

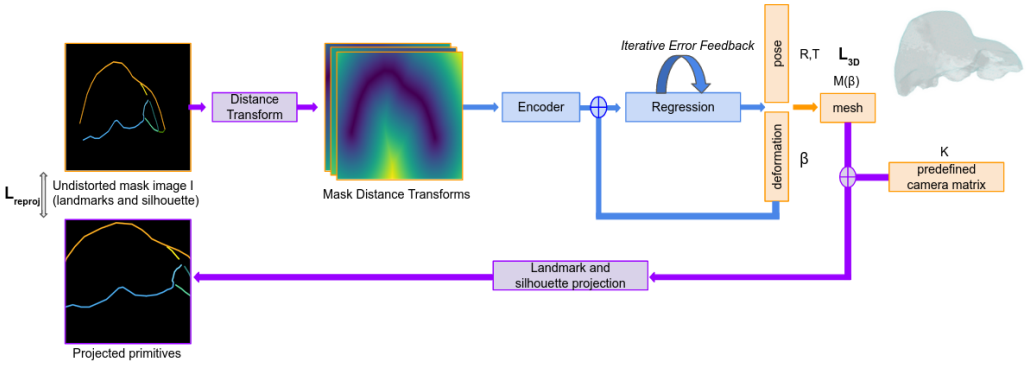


Figure 2: Proposed Liver Mesh Recovery (LMR) framework. Distance maps from image primitive masks are input to an encoder network. Features along with pose and deformation parameters are sent to an iterative regression network to update the parameters, representing a mesh. Reprojection and 3D loss functions are used.

3. Methodology

3.1. Registration Pipeline and the Liver Mesh Recovery Framework

The proposed registration pipeline is shown in figure 1 and the LMR framework in figure 2. LMR uses a ResNet-50 encoder which delivers features from the image primitive masks, represented as distance maps. The features are concatenated with the current pose and deformation parameters. They are used as inputs to the regression network, based on fully connected layers, which iteratively updates the pose R, T and deformation parameters β through IEF (Carreira et al., 2016). We use the same number of layers and neurons in the encoder and regressor networks as in HMR (Kanazawa et al., 2018), except for the input and output layers of the regressor, adapted to the number of deformation parameters. LMR outputs the registered liver mesh, in other words, a mesh properly deformed and expressed in camera coordinates, whose reprojection matches the liver observed in the surgical image.

LMR differs from HMR in four main ways. 1) LMR uses a patient-specific biomechanical liver deformation model, in contrast to an articulated body joint model combined with a generic shape model in HMR. 2) LMR uses the pinhole camera model, which reduces

depth ambiguities, in contrast to the scaled-orthographic model in LMR. 3) LMR is trained patient-specific from semi-synthetic data, as the amount of annotated data in MILR is extremely limited and the real patient-specific data only becomes available at the time of surgery, in contrast to a person-generic model in LMR trained from massive datasets. 4) LMR uses supervised training, making the adversarial discriminator from HMR unnecessary.

3.2. Preoperative Stage

The preoperative steps reconstruct the preoperative 3D model, which is used to synthesise images and train the LMR network.

Step 1: preoperative 3D model reconstruction and landmark annotation. The preoperative volume is segmented, producing a surface mesh, which is upgraded to a volume by constrained Delaunay tetrahedralisation. This yields a volumetric mesh retaining the surface vertices, whose n vertex coordinates, with $\mathcal{O}(n) = 10000$, are stacked into the column vector $\mu^\top = [\mu_1^\top, \cdots, \mu_n^\top] \in \mathbb{R}^{3n}$. The tumours and vena cava are marked as inner regions of the volumetric mesh and the landmarks annotated on the surface. This step is done semi-automatically using standard software including 3D Slicer and Tetgen.

Step 2: deformable liver shape modeling In MILR, the liver mainly undergoes loads from surgical instruments and breathing. Breathing was shown to have a limited influence on the deformation compared to mobilisation (Özgür et al., 2018). We use Finite Element Analysis (FEA) to model the load incurred by a surgical instrument. We apply nodal forces on random vertices of the liver surface. We exclude vertices of the vena cava, which we use as boundary condition, as it forms the main structure maintaining the liver (Flament et al., 1982). We use the Ogden constitutive model with generic liver parameters (Pellicer-Valero et al., 2020). We implement the simulations automatically with the FEBio software (Maas et al., 2012). We preserve the mesh connectivity and thus represent each shape by its vertices in $\mathbb{R}^{3n}$. Upon using forces of random orientation and a load curve of 10 linear steps to reach the magnitude of 0.5 N, we generate $k = 5000$ simulated shapes $x_1, \ldots, x_k$.

With the aim of estimating a deformable shape model with reduced number of parameters, we follow the POD framework (Sifakis and Barbic, 2012). This uses a Truncated SVD of the synthetic shapes $x = [x_1, \cdots, x_k]^\top \in \mathbb{R}^{3n \times k}$ to construct principal components ϕ. We keep the leading m components that allow reconstructing 99% of the data within 5 mm or a minimum of 5 components. A deformed shape is thus represented by its vertex coordinates $\hat{x}_i$, obtained from its m deformation coefficients $\beta_i = [\beta_{i,1}, \cdots, \beta_{i,m}]^\top$ as $\hat{x}_i = \mu + \sum_{j=1}^{m} \beta_{i,j}\phi_j$. Appendix figure 6 illustrates the range of the deformations induced by each model component for one patient.

Step 3: deformation sampling and mask generation. We complete the k FEA simulated shapes to form $l \gg k$, with $l = 40000$ shapes to serve as training dataset. The extra shapes are sampled from a normal distribution using as standard deviation $\frac{2}{3}$ of the one from the initial k coefficients. This guarantees that 99% of the complete set of shapes is within twice the initial standard deviation. We fix the camera intrinsics to a default value, estimated from the laparoscope used for the first patient in our experiments. The next section explains how we handle the actual camera intrinsics, only available in the intraoperative stage. We generate l camera poses by composing a default typical pose with

a random pose perturbation sampled in $SE(3)$. The default typical pose puts the liver in a frontal view where its anterior ridge, ligament and silhouette are mostly visible by the camera. Specifically, it uses X, Y, Z rotation angles of $-65, 45, -45$ degrees and X, Y, Z translations of $0, 0, 175$ mm. The random pose perturbation uses a uniform distribution on the X, Y, Z rotation angles within $50, 30, 40$ degrees and on the X, Y, Z translations of $40, 40, 100$ mm. We use PyTorch3D with its default coordinate system. Appendix figure 5 illustrates the above ranges for one of the patients from our experiments. We project the l shapes using a z-buffer to handle visibility and obtain the image primitives as contours. The dataset has up to $l = 40000$ pairs of shapes and image primitives, from which samples for which fewer than two primitives are visible are removed. We split the dataset in 80% training, 10% validation and 10% test. We convert the primitive contours into segmentation masks, which we randomly perturb to emulate the typical error of automatic detection. Concretely, we simulate missing and thick parts by randomly creating 0 to 3 holes up to 25% of the contour length each. We apply dilation using a disk kernel from 0 up to 10 px side. We finally transform the masks to distance maps, using the Euclidean Distance Transform (Rosenfeld and Pfaltz, 1966) which we normalise to $[-1, 1]$, with 1 being the distance associated with the image diagonal.

Step 4: LMR training. We use a loss with three terms. First, the reprojection loss term, which is the Mean Sum of Distances (MSD) (Li et al., 2005) between the predicted and input primitives. Second, the 3D mesh loss term, which is the Mean Absolute Error (MAE) of the euclidean distances between the predicted and GT mesh vertices. Third, the pose and deformation coefficient loss term, which is the MAE between the normalised predicted and GT pose and deformation coefficients. We use weights of 60, 0.1 and 1 respectively for the three terms. We use the SGD optimiser with 8 samples in the batch, 3 IEF iterations and 50 training epochs. We use the same learning rate for the encoder and regressor, fixed to 1e-3 for the first 45 epochs and then set to 1e-4 with a weight decay of 1e-4.

3.3. Intraoperative Stage

The intraoperative steps extract the primitives from the surgical image and perform registration using LMR. The primitives are extracted automatically by means of segmentation using a COSNet (Lu et al., 2019) specifically trained and validated as in (Labrunie et al., 2022) from 67 patients and 1373 images, from which 100 pairs per patient were constructed to serve co-attention. As our input is a single image, we disable the co-attention module at inference by using the input image twice. We ran an evaluation on the 46 images from the dataset (Rabbani et al., 2021) using the manual primitives. The mean MSD for the baseline (Labrunie et al., 2022) and for the COSNet are 51 px and 40 px, showing an improvement.

The camera is intrinsically calibrated when surgery starts, giving K_{actual}. The parameters were unknown preoperatively; recall that default intrinsics K_{default} were used instead. LMR was thus trained to handle images with different intrinsics than the ones corresponding to the actual surgical image. We cope with this difference by adapting the primitive segmentation masks prior to their use by LMR. Specifically, this is done by applying a 2D affine warp represented by $A \in \mathbb{R}^{3\times3}$, embedding both sets of intrinsics as $A = K_{\text{default}} K_{\text{actual}}^{-1}$, as shown in (Fuentes-Jimenez et al., 2022).

4. Experimental Results

We tested our registration framework with a publicly available clinical dataset[1] designed to evaluate registration accuracy in MILR. This dataset consists of four in-vivo patient cases along with annotated preoperative and intraoperative images, camera calibration, and the facility to measure the predicted tumour Target Registration Error (TRE) with GT obtained by Laparoscopic Ultrasound (LUS) in controlled conditions (Rabbani et al., 2021). We adjusted the landmark annotations of the preoperative 3D model and the images to match the split-ridge model (Labrunie et al., 2022). We used barycentric coordinates to resample the model landmarks to reach four landmark points per mm. We trained LMR from simulations obtained with maximal force magnitudes of 0.5N and 5N. The two results are labelled as OLA (Ogden - Low Amplitude) and OHA (Ogden - High Amplitude). We remove the first 6 deformation components in OHA, which essentially represent rigid displacements and a deformation induced by the fixed vena cava. We conducted an evaluation on semi-synthetic and real data.

4.1. Semi-synthetic Dataset

We use the test dataset generated along with the LMR training data. We measure the Mean Absolute Error (MAE) of the euclidean distances between the estimated and GT mesh vertices, see table 1. The average MAE in each patient is lower than or close to 1 cm. 95% of the test samples have an MAE within $[18, 27]$ mm. OLA and OHA are on par, with differences statistically insignificant. Appendix figure 8 illustrates results for Patient 1.

MAE (mm)		Patient 1	Patient 2	Patient 3	Patient 4	Average
Average	OLA	7.5	6.1	11.4	9.0	8.5
	OHA	9.1	9.8	12.5	8.7	10.0
$\eta_{.95}$	OLA	21.4	18.0	26.9	23.3	22.4
	OHA	22.3	21.2	23.5	20.4	21.9

Table 1: MAE average and 95th percentile on the test dataset for each patient.

4.2. Real Dataset

We measure the tumour TRE of the proposed LMR for the real dataset (Rabbani et al., 2021). We compare with the following three baselines. M-Pose is the rigid pose manually given by an expert. A-Pose is a rigid pose estimated from the primitives (Labrunie et al., 2022). Opt-B is a deformation estimated by refining the result of A-Pose (Labrunie et al., 2022), using a Neo-Hookean constitutive deformation model. The methods are run from manual or automatic image annotations. The results are given in table 2. Opt-B obtains the best performance in all patients for manual annotations, with a TRE within $[9, 15]$ mm, except for Patient 2 where a strong deformation induced by the ultrasound probe leads to over 6 cm TRE. LMR mainly obtains TRE lower than 2 cm for OLA with a smaller standard deviation, and higher TRE and standard deviation for OHA. It also obtains a high TRE for Patient 2 but outperforms Opt-B. The performance of both methods degrade for automatic

1. http://igt.ip.uca.fr/~ab/code_and_datasets/datasets/llr_reg_evaluation_by_lus

TRE (mm)		Patient 1		Patient 2		Patient 3		Patient 4	
Methods		A	M	A	M	A	M	A	M
M-Pose		$15.1^{\pm 2.3}$		$35.5^{\pm 12.2}$		$30.5^{\pm 5.2}$		$16.3^{\pm 2.3}$	
A-Pose		$30.0^{\pm 10.8}$	$9.8^{\pm 3.0}$	$86.9^{\pm 14.7}$	$63.0^{\pm 11.4}$	$29.8^{\pm 60.2}$	$9.7^{\pm 2.3}$	$20.5^{\pm 1.0}$	$17.8^{\pm 0.8}$
Opt-B		$29.3^{\pm 10.6}$	$10.3^{\pm 2.8}$	$86.9^{\pm 14.7}$	$63.0^{\pm 11.4}$	$30.0^{\pm 60.1}$	$9.5^{\pm 2.1}$	$19.8^{\pm 1.0}$	$14.7^{\pm 1.2}$
LMR	OLA	$17.9^{\pm 3.7}$	$17.4^{\pm 4.2}$	$45.6^{\pm 14.7}$	$53.8^{\pm 8.0}$	$46.9^{\pm 7.1}$	$17.6^{\pm 0.6}$	$22.7^{\pm 9.1}$	$17.0^{\pm 0.8}$
	OHA	$16.1^{\pm 1.6}$	$21.6^{\pm 5.4}$	$44.7^{\pm 9.9}$	$53.2^{\pm 11.7}$	$53.1^{\pm 9.5}$	$28.3^{\pm 0.7}$	$23.0^{\pm 5.7}$	$30.5^{\pm 1.7}$

Table 2: Average tumour TRE and standard deviation ($\pm$) for the real data. Columns A and M respectively stand for automatic and manual image annotations.

image annotation, owing to segmentation errors (see appendix figure 10). Appendix figure 9 shows the detected and reprojected primitives for OLA. The observed discrepancy suggests that the methods would benefit from higher deformation variability in the training data.

All experiments were run on a medium-end Nvidia GeForce RTX 2080 GPU. The inference time was 5 ± 3 s for A-Pose, 43 ± 47 s for the optimisation in Opt-B and 3.4 ± 0.2 ms for LMR inference. Primitive detection with COSNet takes 9.3 ± 0.2 ms. This means that our pipeline allows real-time processing, with at least 16 ± 2 fps as obtained with our unoptimised implementation. FEA simulation and mask generation took 4 hours to run and LMR 16 hours to train per patient. This can be reduced by using a high-end GPU, but is already compatible with the typical delay between preoperative scanning and surgery.

5. Conclusion

We have proposed LMR, a novel patient-specific learning-based registration method in MILR. LMR directly regresses the registration parameters from primitives detected in the intraoperative image. In contrast to the state-of-the-art optimisation-based approaches, LMR transfers most of the computation to the preoperative stage, where it is trained from physics-based simulations. Inference takes a few forward passes -we use three passes- through a neural model. It is initialisation-free and fast, offering a speed up of five orders of magnitude compared to the baseline optimisation-based method. Our experimental results suggest that LMR obtains competitive results in real clinical cases, achieving a tumour TRE lower than 2 cm, except for a large deformation case defeating all existing methods. Our pipeline would benefit from three types of improvements. First, the accuracy of the proposed primitive detector could be improved by training on a much larger dataset. Second, the physics-based simulation conditions are overly simplistic to represent some of the clinical cases well and would benefit from additional loads such as breathing and the pneumoperitoneum. Third, the physics-based simulation and LMR training start from scratch on each patient. Given a larger MILR dataset, a patient-generic model could be trained once and then used as initial model for patient-specific fine-tuning.

Acknowledgments

Acknowledgments withheld.

References

Mourad Adballah, Yamid Espinel, Lilian Calvet, Bruno Pereira, Bertrand Le Roy, Adrien Bartoli, and Emmanuel Buc. Augmented reality in laparoscopic liver resection evaluated on an ex-vivo animal model with pseudo-tumours. *Surgical Endoscopy*, 36(1):833–843, 2022.

Joao Carreira, Pulkit Agrawal, Katerina Fragkiadaki, and Jitendra Malik. Human pose estimation with iterative error feedback. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4733–4742, 2016.

Yamid Espinel, Erol Özgür, Lilian Calvet, Bertrand Le Roy, Emmanuel Buc, and Adrien Bartoli. Combining visual cues with interactions for 3d–2d registration in liver laparoscopy. *Annals of Biomedical Engineering*, 48(6):1712–1727, 2020.

JB Flament, JF Delattre, and G Hidden. The mechanisms responsible for stabilising the liver. *Anatomia Clinica*, 4(2):125–135, 1982.

David Fuentes-Jimenez, Daniel Pizarro, David Casillas-Pérez, Toby Collins, and Adrien Bartoli. Deep shape-from-template: Single-image quasi-isometric deformable registration and reconstruction. *Image and Vision Computing*, 127:104531, 2022.

Peng Guan, Alexander Weiss, Alexandru O Balan, and Michael J Black. Estimating human shape and pose from a single image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1381–1388. IEEE, 2009.

Angjoo Kanazawa, Michael J Black, David W Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7122–7131, 2018.

Bongjin Koo, Erol Özgür, Bertrand Le Roy, Emmanuel Buc, and Adrien Bartoli. Deformable registration of a preoperative 3d liver volume to a laparoscopy image using contour and shading cues. In *International conference on medical image computing and computer-assisted intervention*, pages 326–334. Springer, 2017.

Bongjin Koo, Maria R Robu, Moustafa Allam, Micha Pfeiffer, Stephen Thompson, Kurinchi Gurusamy, Brian Davidson, Stefanie Speidel, David Hawkes, Danail Stoyanov, et al. Automatic, global registration in laparoscopic liver surgery. *International Journal of Computer Assisted Radiology and Surgery*, 17(1):167–176, 2022.

M Labrunie, M Ribeiro, F Mourthadhoi, C Tilmant, B Le Roy, E Buc, and A Bartoli. Automatic preoperative 3d model registration in laparoscopic liver resection. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–8, 2022.

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

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6):1–16, 2015.

Xiankai Lu, Wenguan Wang, Chao Ma, Jianbing Shen, Ling Shao, and Fatih Porikli. See more, know more: Unsupervised video object segmentation with co-attention siamese networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3623–3632, 2019.

Steve A Maas, Benjamin J Ellis, Gerard A Ateshian, and Jeffrey A Weiss. Febio: finite elements for biomechanics. *Journal of biomechanical engineering*, 134(1), 2012.

Richard Modrzejewski. *Recalage déformable, jeux de données et protocoles d'évaluation pour la chirurgie mini-invasive abdominale augmentée*. PhD thesis, Université Clermont Auvergne (2017-2020), 2020.

Erol Özgür, Bongjin Koo, Bertrand Le Roy, Emmanuel Buc, and Adrien Bartoli. Preoperative liver registration for augmented monocular laparoscopy using backward–forward biomechanical simulation. *International journal of computer assisted radiology and surgery*, 13(10):1629–1640, 2018.

Oscar J Pellicer-Valero, María José Rupérez, Sandra Martínez-Sanchis, and José D Martín-Guerrero. Real-time biomechanical modeling of the liver using machine learning models trained on finite element method simulations. *Expert Systems with Applications*, 143: 113083, 2020.

Rosalie Plantefeve, Nazim Haouchine, Jean-Pierre Radoux, and Stéphane Cotin. Automatic alignment of pre and intraoperative data using anatomical landmarks for augmented laparoscopic liver surgery. In *International Symposium on Biomedical Simulation*, pages 58–66. Springer, 2014.

N Rabbani, L Calvet, Y Espinel, B Le Roy, M Ribeiro, E Buc, and A Bartoli. 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 & Visualization*, pages 1–10, 2021.

Maria R Robu, João Ramalhinho, Stephen Thompson, Kurinchi Gurusamy, Brian Davidson, David Hawkes, Danail Stoyanov, and Matthew J Clarkson. Global rigid registration of ct to video in laparoscopic liver surgery. *International Journal of Computer Assisted Radiology and Surgery*, 13(6):947–956, 2018.

Azriel Rosenfeld and John L Pfaltz. Sequential operations in digital picture processing. *Journal of the ACM (JACM)*, 13(4):471–494, 1966.

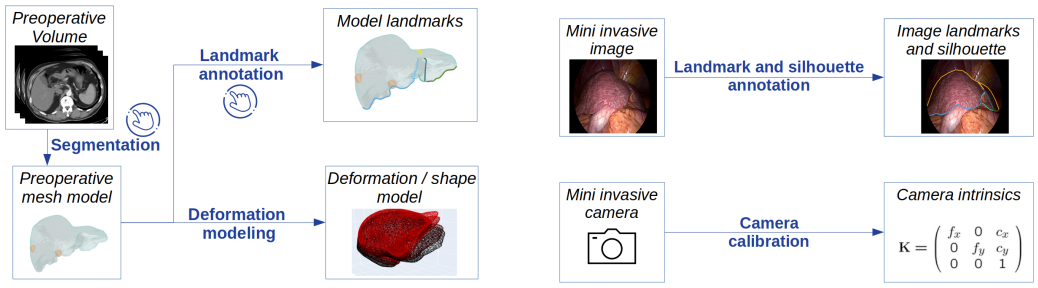
Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.

Eftychios Sifakis and Jernej Barbic. Fem simulation of 3d deformable solids: a practitioner's guide to theory, discretization and model reduction. In *Acm siggraph 2012 courses*, pages 1–50. 2012.

Yating Tian, Hongwen Zhang, Yebin Liu, and Limin Wang. Recovering 3d human mesh from monocular images: A survey. Eprint arXiv:2203.01923, 2022.

Fu-Ping Zhong, Yu-Jun Zhang, Ying Liu, and Shu-Bing Zou. Prognostic impact of surgical margin in patients with hepatocellular carcinoma: a meta-analysis. *Medicine*, 96(37), 2017.

Appendix A. Supplementary figures



(*a*) Preoperative (*b*) Intraoperative

Figure 3: Steps followed to solve registration in MILR, common to both existing optimisation-based methods and the proposed learning-based method.

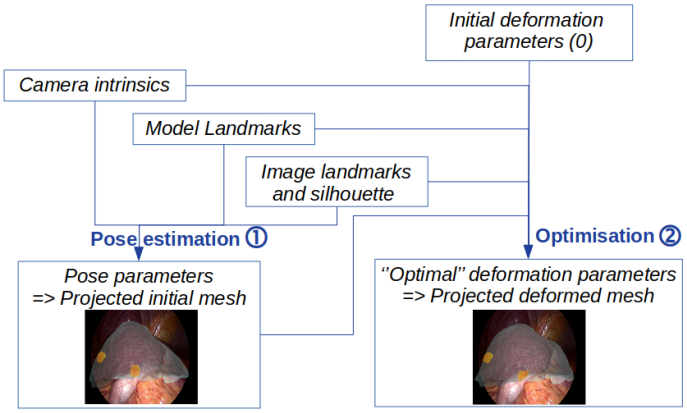
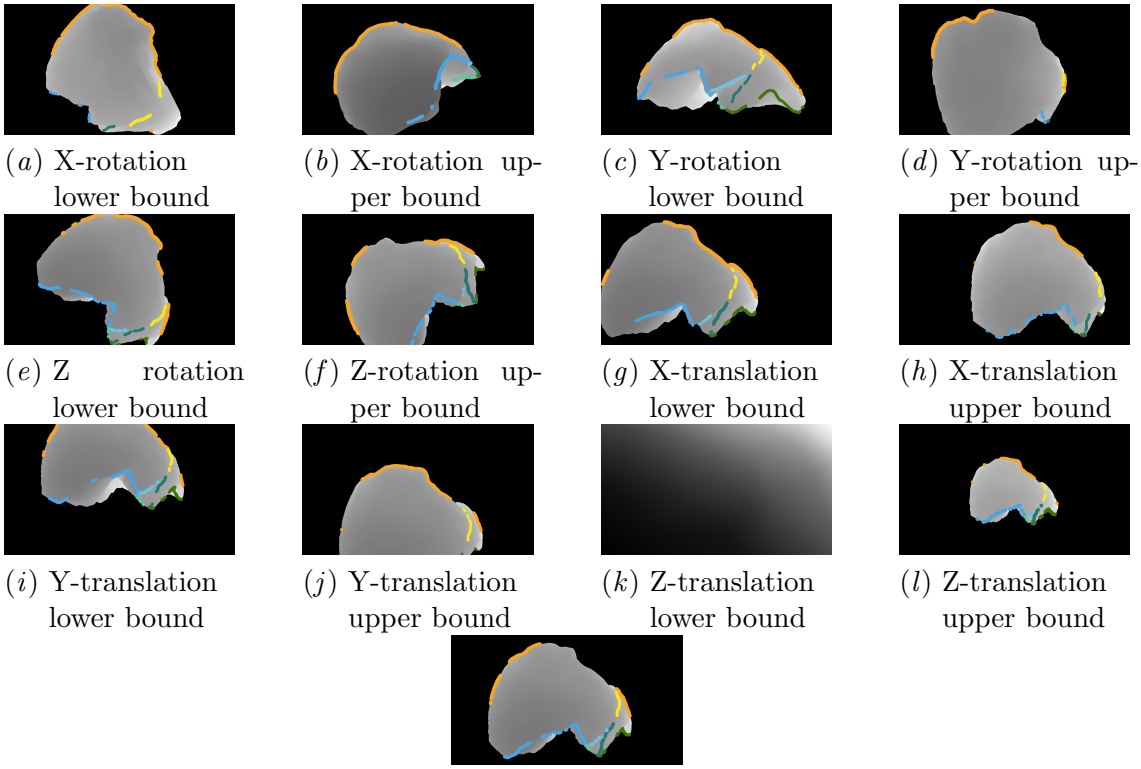


Figure 4: Steps to solve registration in MILR specific to the optimisation-based approach.



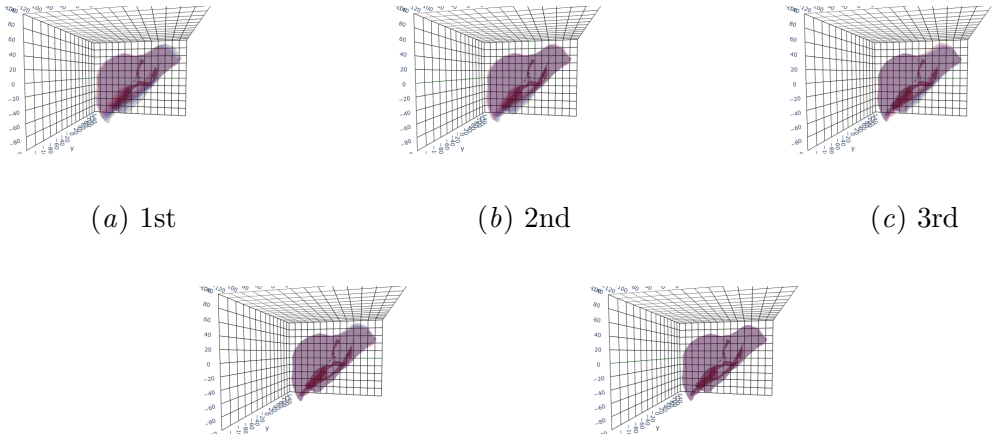
(*a*) X-rotation lower bound (*b*) X-rotation upper bound (*c*) Y-rotation lower bound (*d*) Y-rotation upper bound

(*e*) Z rotation lower bound (*f*) Z-rotation upper bound (*g*) X-translation lower bound (*h*) X-translation upper bound

(*i*) Y-translation lower bound (*j*) Y-translation upper bound (*k*) Z-translation lower bound (*l*) Z-translation upper bound

(*m*) Initial position

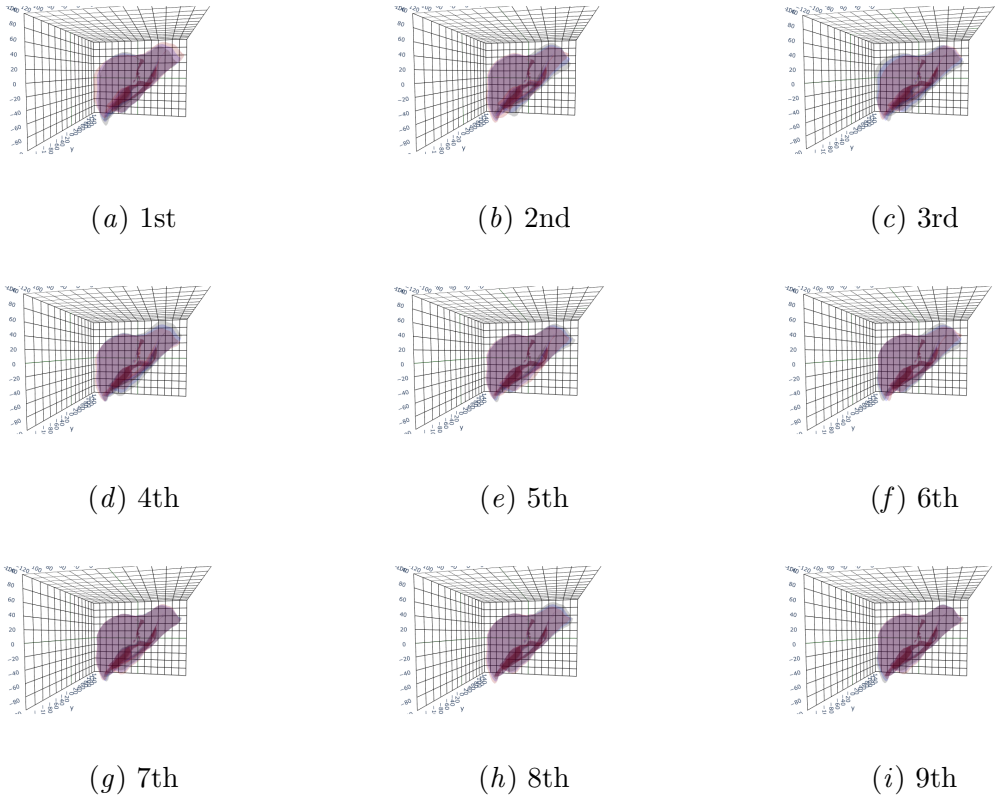
Figure 5: Undeformed liver configurations obtained within the bounds of the pose simulation range where only one parameter is modified compared with the reference configuration (m), for Patient 4. Examples of advanced simulated primitives are superimposed to the depth map and enlarged for visualisation purposes. The silhouette, the ligament junction and the split ridge are respectively in orange, yellow and nuances of green and blue.



(a) 1st (b) 2nd (c) 3rd

(d) 4th (e) 5th

Figure 6: Nomograms of the deformation components for OLA in Patient 1. In blue, the initial mesh, and in red and black the deformed meshes within the bounds of the deformation ranges.



(a) 1st (b) 2nd (c) 3rd

(d) 4th (e) 5th (f) 6th

(g) 7th (h) 8th (i) 9th

Figure 7: Nomograms of the deformation components for OHA in Patient 1. In blue, the initial mesh, and in red and black the deformed meshes within the bounds of the deformation ranges.

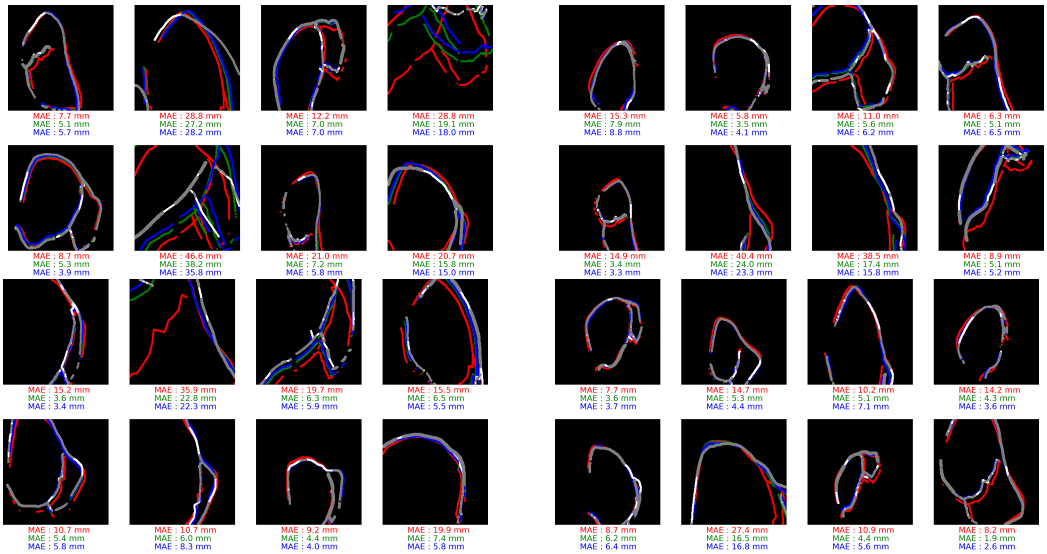
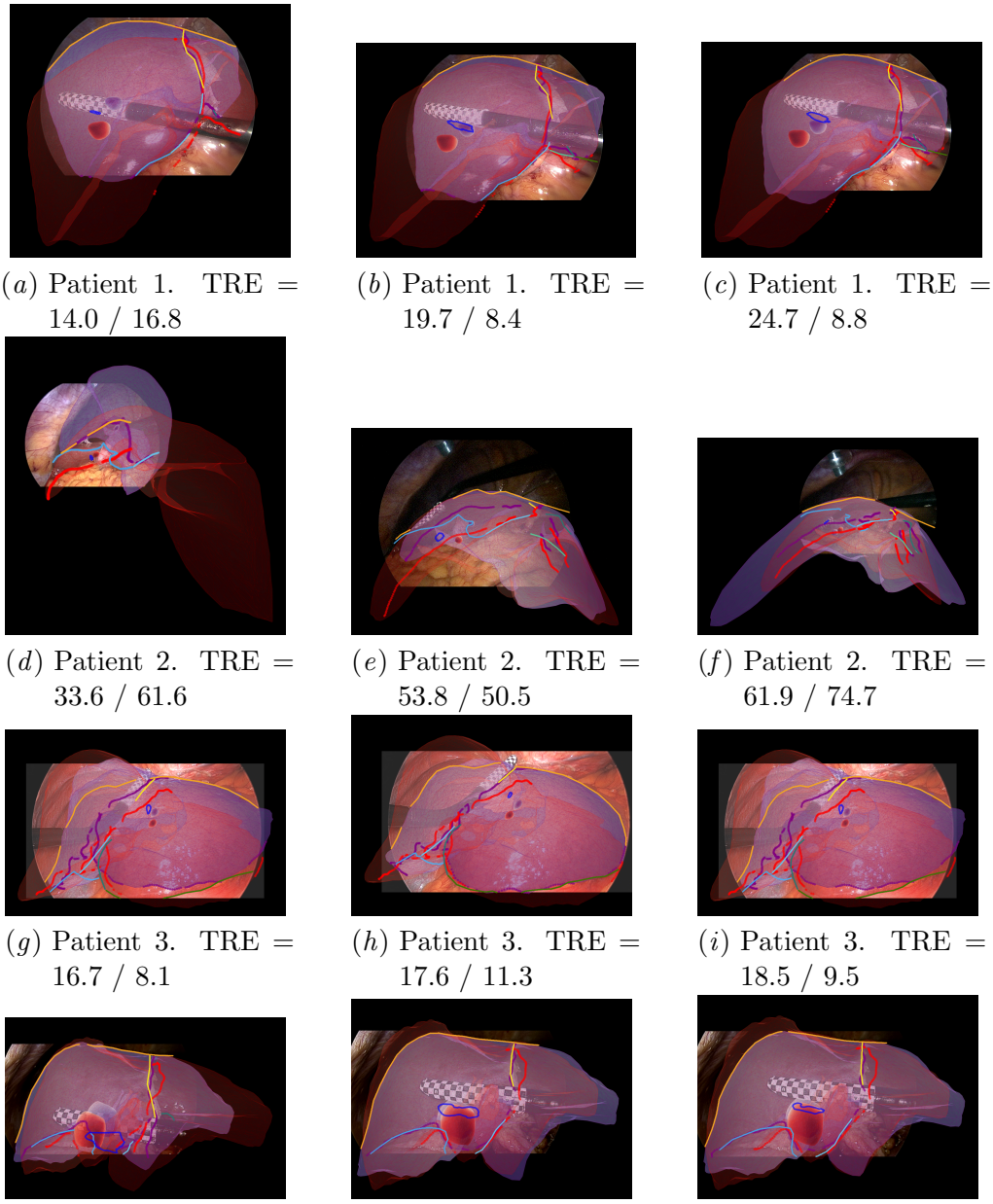


Figure 8: Results on the test dataset for 32 samples of Patient 1, using OLA. In gray, the initial segmentation mask, in white the projected GT primitives and in red, green and blue the respective projected results of each IEF iteration with associated MAE values.



(*a*) Patient 1. TRE = 14.0 / 16.8

(*b*) Patient 1. TRE = 19.7 / 8.4

(*c*) Patient 1. TRE = 24.7 / 8.8

(*d*) Patient 2. TRE = 33.6 / 61.6

(*e*) Patient 2. TRE = 53.8 / 50.5

(*f*) Patient 2. TRE = 61.9 / 74.7

(*g*) Patient 3. TRE = 16.7 / 8.1

(*h*) Patient 3. TRE = 17.6 / 11.3

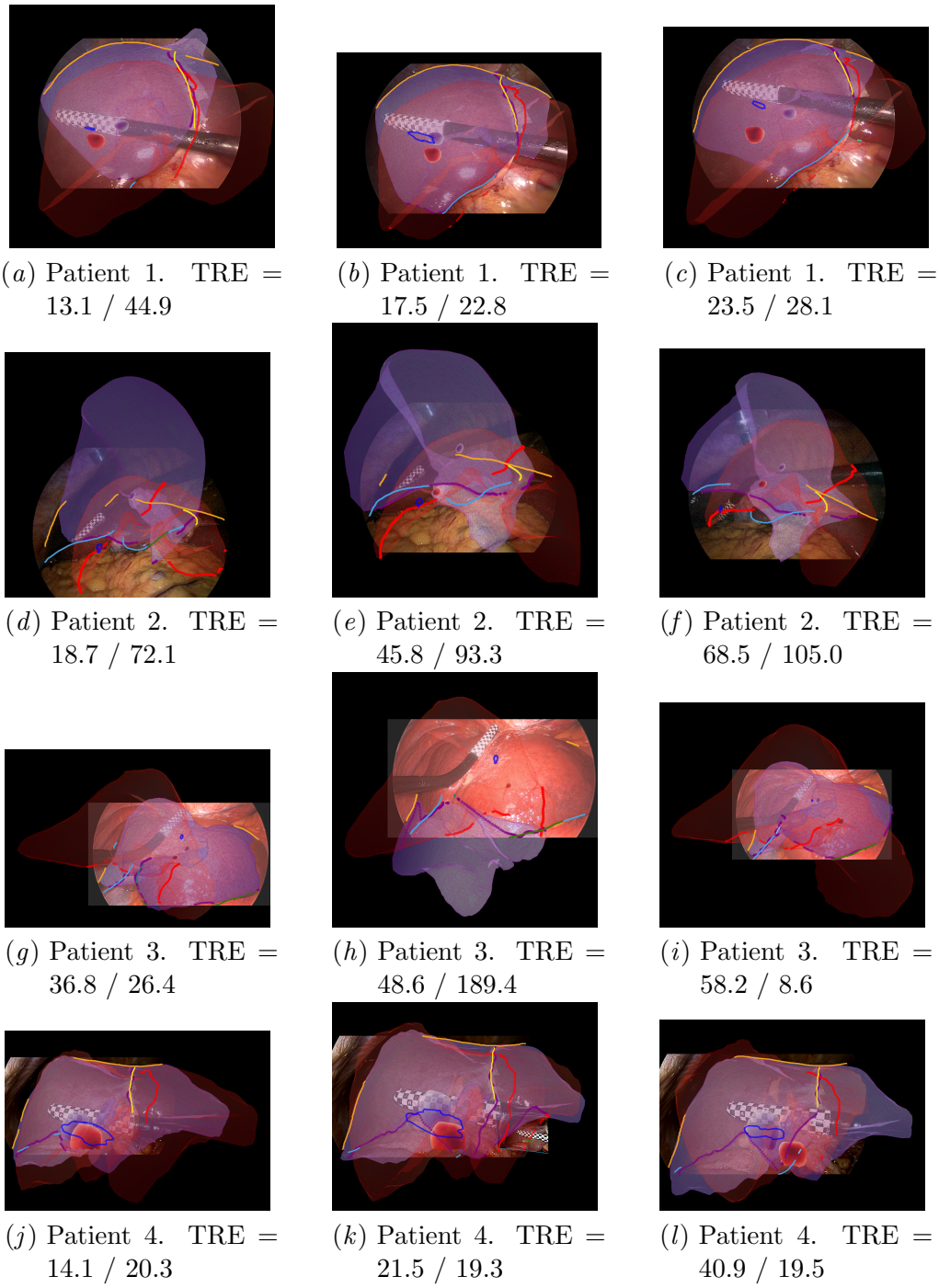
(*i*) Patient 3. TRE = 18.5 / 9.5

(*j*) Patient 4. TRE = 15.7 / 16.9

(*k*) Patient 4. TRE = 17.0 / 13.4

(*l*) Patient 4. TRE = 18.2 / 14.6

Figure 9: Comparative results between the Opt-B baseline and the proposed LMR-OLA method, for manual annotation of the primitives. Three images of each patient, selected as the minimal (left) and maximal (right) TRE, and the closest TRE to the mean (middle) of the LMR-OLA method. TREs given in the captions are respectively for LMR-OLA / Opt-B. Opt-B results are in purple and the LMR ones are in red. Visible model landmarks are additionally displayed using the same colour code. The projected tumour outline registered from the ultrasound is in blue.



(*a*) Patient 1. TRE = 13.1 / 44.9

(*b*) Patient 1. TRE = 17.5 / 22.8

(*c*) Patient 1. TRE = 23.5 / 28.1

(*d*) Patient 2. TRE = 18.7 / 72.1

(*e*) Patient 2. TRE = 45.8 / 93.3

(*f*) Patient 2. TRE = 68.5 / 105.0

(*g*) Patient 3. TRE = 36.8 / 26.4

(*h*) Patient 3. TRE = 48.6 / 189.4

(*i*) Patient 3. TRE = 58.2 / 8.6

(*j*) Patient 4. TRE = 14.1 / 20.3

(*k*) Patient 4. TRE = 21.5 / 19.3

(*l*) Patient 4. TRE = 40.9 / 19.5

Figure 10: Comparative results between the Opt-B baseline and the proposed LMR-OLA method, however for automatic annotation of the primitives. The same structure and colour code as Figure 9 are used.