Sara Schwab*, Lorena Krames, and Werner Nahm

Influencing Factors on the Registration Accuracy of a Learned Feature Descriptor in Laparoscopic Liver Surgery

https://doi.org/10.1515/cdbme-2024-2139

Abstract: In laparoscopic liver surgery, image-guided navigation systems provide crucial support to surgeons by supplying information about tumor and vessel positions. For this purpose, these information from a preoperative CT or MRI scan is overlaid onto the laparoscopic video. One option is performing a registration of preoperative 3D data and 3D reconstructed laparoscopic data. A robust registration is challenging due to factors like limited field of view, liver deformations, and 3D reconstruction errors. Since in reality various influencing factors always intertwine, it is crucial to analyze their combined effects. This paper assesses registration accuracy under various synthetically simulated influences: patch size, spatial displacement, Gaussian deformations, holes, and downsampling. The objective is to provide insights into the required quality of the intraoperative 3D surface patches. LiverMatch serves as the feature descriptor, and registration employs the RANSAC algorithm. The results of this paper show that ensuring a large field of view of at least 15-20% of the liver surface is necessary, allowing tolerance for less accurate depth estimation.

Keywords: Laparoscopic liver surgery, 3D-3D registration, 3D local feature descriptor, synthetic data

1 Introduction

Laparoscopic liver surgery (LLS) offers patients benefits such as reduced pain and shorter hospital stays compared to open surgery [1]. However, LLS presents challenges like limited field of view and lack of haptic feedback. Visualizing tumors and vessels derived from preoperative data as overlay on the laparoscopic screen could support surgeons [2]. This requires the registration of preoperative information from a 3D model segmented from CT or MRI scans with the laparoscopic video. The registration is challenging due to liver surface features' lack of distinctiveness and influencing factors like deformations. One option is a 3D-3D registration, which requires the 3D reconstruction of the laparoscopic images. Two important steps of a state-of-the-art registration pipeline are the feature descriptor and matcher. The feature descriptor encodes the pointwise features of both, the preoperative and the intraoperative data. Afterwards, the feature matcher finds correspondences between the two data sets based on the feature description and aligns them accordingly.

Various approaches have been proposed to solve the registration problem. Pfeiffer et al. suggested a convolutional neural network for both feature generation and non-rigid matching [3]. However, this approach requires a prior rigid alignment of the point sets. Robu et al. used the triple orthogonal local depth images (TOLDI) as handcrafted feature descriptor combined with the random sample consensus (RANSAC) as feature matcher [4]. However, handcrafted descriptors often struggle with realistic influencing factors e.g., spatial displacement and deformations [5]. Novel approaches like LiverMatch introduced by Yang et al., use learned feature descriptors to improve the registration accuracy [6].

Typically, registration performance is analyzed using a data set with fixed influencing factor ranges or by analyzing influencing factors separately. However, in real-world scenarios, multiple influencing factors occur simultaneously, making it crucial to assess their combined influence and interactions. This analysis aids in identifying high influencing factors and determining limitations where improvement with other techniques is necessary to ensure adequate registration performance.

The purpose of this paper is to analyze the impact on registration accuracy under various combinations of influencing factors, in order to gain a deeper understanding of their interactions and limitations. The emphasis lies in quantifying the impact of influencing factors for the first time by using a simulation environment and drawing a conclusive statement regarding the required intraoperative data quality.

2 Materials and Methods

2.1 Dataset Preparation

For the preoperative data, the 3D-IRCADb-01 dataset consisting of 20 3D human liver models segmented from CT scans was used [7]. The liver models 1, 6, and 18 were used for testing and 15, 17, and 20 for validating the learned feature

***Corresponding author: Sara Schwab,** Institute of Biomedical Engineering (IBT), Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131 Karlsruhe, Germany, e-mail: publications@ibt.kit.edu
Lorena Krames, Werner Nahm, IBT, KIT, 76131 Karlsruhe, Germany

descriptor. The remaining models were used for training. Synthetic intraoperative data were generated using Krames et al.'s data generation pipeline due to the unavailability of real 3D laparoscopic liver data. In a first step, a surface patch is cropped from the preoperative point cloud, its size as percentage of the whole model size is specified with the parameter f_{ps}. Afterwards, various influencing factors i.e., spatial displacement, Gaussian deformation (amplitude defined by f_{gd}), and downsampling, with adjustable degrees are applied to the surface patch. Lastly, the surface patch is rigidly transformed. Additionally, a ground truth surface patch is generated for evaluation. For further information we refer to [8]. For this work, the pipeline was modified, including an additional influencing factor called holes, and adjusting the implementation of spatial displacement, downsampling, and patch transformation.

Implementation of Holes: This influencing factor simulates parts of the intraoperative data, where the depth estimation failed due to disadvantageous light conditions such as reflections or dark spots. These parts are implemented as holes with a defined size (percentage of the patch f_{hs}) and number (f_{hn}), and are evenly distributed on the patch. **Adjustment of spatial displacement:** Here, point coordinates (x,y,z) changed bi-directionally to avoid a global shift of the patch. Thus, a spatial displacement (f_{sd}) of 5 mm described coordinate changes spanning from -2.5 mm to 2.5 mm instead of 0 mm to 5 mm. **Adjustment of downsampling:** In this work, the point density of the intraoperative surface patch varied, instead of the point density of the preoperative point cloud. The downsampling factor (f_{ds}) here described the ratio of original to remaining points after downsampling. At $f_{ds} = 1$, no downsampling occurred; at $f_{ds} = 0.8$, only 80% of all points of the patch were used. **Adjustment of transformation:** Instead of a fixed rotation and translation, intraoperative patches were randomly rotated. Additionally, all point clouds were centered, with preoperative point clouds sharing a consistent orientation within the coordinate system.

Figure 1 shows examples for all influencing factors (i.e., patch size, spatial displacement, Gaussian deformation, downsampling, holes) of the adjusted data generation pipeline.

2.2 Registration Pipeline

The registration pipeline consists of a features descriptor, a feature matcher and the evaluating registration accuracy.

Learned Feature Descriptor: Learned feature descriptors are based on the usage of artificial neural networks to generate the feature encoding. In this study, LiverMatch specifically tailored for 3D-3D laparoscopic liver registration, was used [6]. LiverMatch utilizes preoperative and intraoperative point clouds to generate feature descriptions for each point, along with visibility scores and matches. For further information regarding the LiverMatch we refer to [6]. Training followed the original implementation's hyperparameters, except for the maximum number of epochs set to 40 with an early stopping algorithm to cease training if no improvement occurred for 5 epochs. LiverMatch was trained with 14,000 examples and validated with 3,000 examples, each containing 125 patches per liver model. Various values (10%, 15%, 20% and 25%) for f_{ps} were used, and 1 and 0.75 for f_{ds}. All eight combinations of these parameters were applied to each patch. The parameter f_{sd} varied from 0 mm to 25 mm in 5 mm increments, and f_{gd} included degrees of 0 mm, 5 mm, 10 mm, 20 mm, and 30 mm. Random combinations of these degrees were added to each patch, ensuring an even distribution of each factor and degree for each liver model.

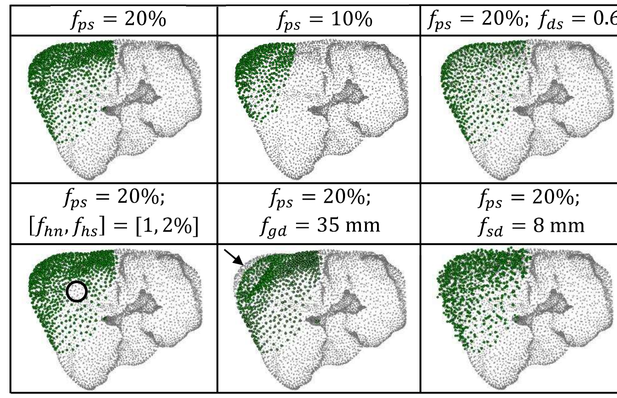

Fig. 1: Examples of intraoperative patches with different degrees of the influencing factors i.e., patch size (f_{ps}), holes ($[f_{hn}, f_{hs}]$), downsampling factor (f_{ds}), spatial displacement (f_{sd}) and Gaussian deformation (f_{gd}. The ground truth point cloud (patch) is shown in green and the preoperative point cloud in gray.

Feature Matcher: RANSAC with default settings, implemented in the Open3D package, was used [9, 10]. It estimates correspondences and calculates a rigid transformation based on the two input point clouds and the computed features.

Evaluation: Registration accuracy is assessed using the root mean square error (RMSE). First, the registered ground truth patch is computed by applying the transformation matrix from RANSAC to the points. Afterwards, RMSE is calculated between the points of the registered ground truth patch and the intraoperative surface patch. To evaluate registration accuracy in LLS, comparison with two technical and clinical motivated threshold values is conducted: 5.38 mm, reflecting average registration accuracy observed in LLS [11], and 10 mm, derived from the recommended safety margin for tumor resection [12].

2.3 Experiment

To analyze the impact of different influencing factors on registration accuracy, we created a data set with varied combinations and degrees of the influencing factors, carefully selected to closely mimic real-world scenarios. The data set comprises 2,700 intraoperative patches, with 50 patches per liver model. Parameters f_{ps} ranged from 20%, 15% to 10%, and f_{ds} from 1, 0.8, to 0.6. Additionally, either no holes or holes with $f_{hn} = 5$ and $f_{hs} = 2\%$ were included. All combinations of these factors were applied to each patch, along with random combinations of f_{sd} (ranging from 1 mm to 15 mm in increments of 3.5 mm) and f_{gd} (ranging from 0 mm to 35 mm in increments of 8.75 mm). To ensure balanced representation, all degrees of f_{sd} and f_{gd} occurred evenly in the data set. Given the stochastic nature of RANSAC, the algorithm was averaged over 30 iterations for each patch. The RMSE for evaluation represents the mean value across these repetitions.

3 Results

To better understand the impact of various influencing factors on registration performance, Table 1 presents the percentage increase in mean RMSE when modifying one influencing factor. The following key observations can be made from the data: While holes had a minor impact, reducing patch size by 5% led to an approximately 80% increase in mean RMSE. Downsampling exhibited a greater percentage increase when reduced from 0.8 to 0.6 compared to 1 to 0.8. No consistent trend was observed for spatial displacement or Gaussian deformation. However, the percentage increase in spatial displacement was highest for the smallest degrees, while for Gaussian deformation, it was highest for the largest. While increasing spatial displacement by 3.5 mm had a large influence in some cases, increasing Gaussian deformation by 8.75 mm had only a minor impact.

Figure 2 provides additional insights into the effects of various combinations of spatial displacement, downsampling, and patch size. The data illustrates: With larger patch sizes, more mean values fell below the 10 mm threshold; notably, at a 10% patch size, only one combination achieved a mean below 5.38 mm. Increasing f_{sd} and decreasing f_{ds} typically resulted in decreased mean RMSE in most scenarios.

Tab. 1: Percentage increase in the mean value of the RMSE when varying only the degree of the specific influencing factors.

Change of influencing factor	Percentage increase of mean RMSE
f_{ps}: 20% to 15%	79%
f_{ps}: 15% to 10%	85%
f_{ds}: 1 to 0.8	22%
f_{ds}: 0.8 to 0.6	67%
f_{sd}: 1 mm to 4.5 mm	59%
f_{sd}: 4.5 mm to 8 mm	20%
f_{sd}: 8 mm to 11.5 mm	48%
f_{sd}: 11.5 mm to 15 mm	22%
f_{gd}: 0 mm to 8.75 mm	14%
f_{gd}: 8.75 mm to 17.5 mm	19%
f_{gd}: 17.5 mm to 26.25 mm	7%
f_{gd}: 26.25 mm to 35 mm	26%
$[f_{hn}, f_{hs}]$: [0, 0%] to [5, 2%]	21%

4 Discussion

The aim of this study was to analyze how various influencing factors affect registration accuracy by examining different combinations of these factors. Adding holes with f_{hn} = 5 and f_{hs} = 2% (removing 10% of patch points) minimally affects registration accuracy. The remaining points still offer sufficient information for similarly registration accuracy. Regarding downsampling, the impact of reducing from 0.8 to 0.6 exceeds that from 1 to 0.8. Possibly this observation is a result of a 40% reduction in points, leading to an insufficient point count for registration. Additionally, the feature descriptor's training with f_{ds} values of 1 and 0.75 could limit its capability to handle smaller f_{ds} values like 0.6. Overall, a minimum number of points seems crucial for sufficient information, as emphasized by the high impact of patch size.

An increase in Gaussian deformation results in a percentage increase of over 20% only when changing from 26.25 mm to 35 mm. Consequently, small Gaussian deformations seem to have minimal influence, likely because it affects only a part of the patch [8].

As spatial displacement increased by 3.5 mm, high variability was observed in percentage increases, likely due to the random selection of the exact point coordinate changes. Nonetheless, overall, spatial displacement had a considerable impact, likely because it alters all points' positions. Moreover, it simulates depth estimation, emphasizing the need for precise accuracy to ensure high registration accuracy. Hattab et al. also demonstrated that removing noise or spatial displacement improves registration [13]. However, when combined with patch size, the influence of the latter dominates. With a larger patch size, larger degrees of the other factors can be tolerated, including potentially inaccurate depth estimation. For example, with a patch size of 10%, only at $f_{ds} = 1$ and a highly accurate depth estimation of $f_{sd} = 1$ mm, a registration accuracy below 5.38 mm was achieved, whereas with a patch size of 20%, this applies even at $f_{ds} = 1$ and $f_{sd} = 8$ mm. This emphasizes the need for a larger patch size to obtain more information. Moreover, this confirms the inherent challenges in LLS registration due to the small field of view, which varies approximately between 10% and 20% [3, 4], and limited liver features [3]. Methods to increase patch size, like image or point stitching, should be implemented prior to the registration pipeline.

It is notably that increasing f_{sd} and decreasing f_{ds} do not consistently decrease registration accuracy. This inconsistency may stem from the random assignment of f_{sd} and f_{gd} combinations to each patch. Consequently, not every patch encompasses all combinations. This results in the observed discrep-

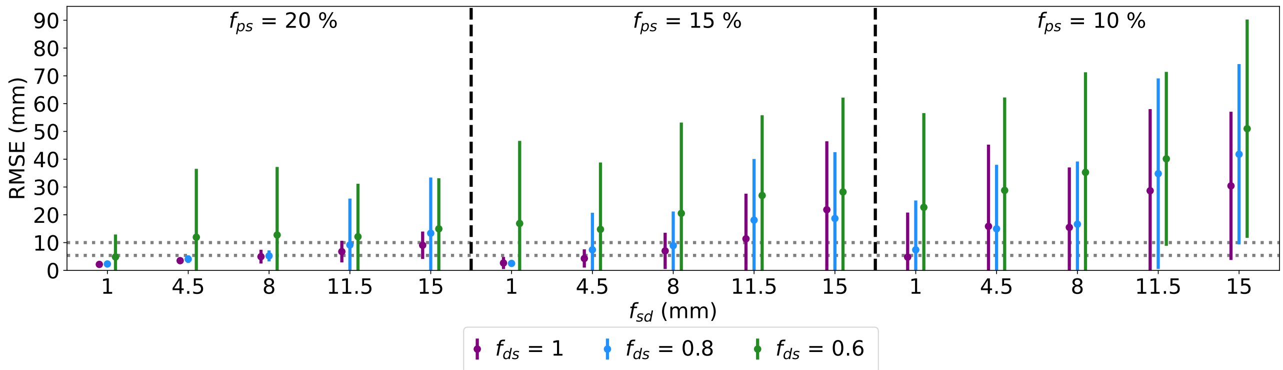

Fig. 2: RMSE (mean and standard deviation) for different combinations of patch size (f_{ps}), spatial displacement (f_{sd}) and downsampling factor (f_{ds}). The horizontal dotted gray lines depict the two thresholds at 5.38 mm and 10 mm.

ancies because the patch's position also influences registration accuracy, as demonstrated in [5].

This study is subject to certain limitations. First, the adjusted data generation pipeline simplifies the real-world LLS setting, e.g., by using Gaussian deformation and spatial displacement, which cannot fully replicate all real deformations and depth estimation errors. Additionally, LiverMatch was trained with a limited data set and degrees of influencing factors. Expanding the data set and incorporating a wider range of factors could likely enhance registration accuracy. Lastly, the results are influenced by the data set, lacking all combinations of influencing factors for every patch.

5 Conclusion

A patch size of at least 15% to 20% is crucial for acceptable registration accuracy, as it allows for larger combined influencing factors. Therefore, methods such as image or point cloud stitching should be applied in the registration pipeline before the feature descriptor to increase field of view. This confirms that the registration challenge in LLS is ill-posed due to the limited field of view and the lack of unique landmarks on the liver surface [3]. While accurate depth estimation is essential, the impact of a large field of view prevails in this context, allowing for the acceptance of greater errors in depth estimation.

Author Statement

This research was funded by Olympus Surgical Technologies Europe. There is no further conflict of interest to declare.

References

[1] Reddy SK, Tsung A, Geller DA. Laparoscopic Liver Resection. World J Surg. 2011; 35:1478–1486.

[2] Cleary K, Peters T. Image-Guided Interventions: Technology Review and Clinical Applications. Annu. Rev. Biomed. Eng. 2010; 12:119–142.

[3] Pfeiffer M, Riediger C, Leger S, Kühn JP, Seppelt D, Hoffmann RT et al. Non-Rigid Volume to Surface Registration Using a Data-Driven Biomechanical Model. MICCAI 2020; 12264:724–734.

[4] Robu MR, Ramalhinho J, Thompson S, Gurusamy K, Davidson B, Hawkes D, et al. Global Rigid Registration of CT to Video in Laparoscopic Liver Surgery. Int J Comput Assist Radiol Surg. 2018 Jun; 13(6):947-956.

[5] Krames L, Suppa P, Nahm W. Does the 3D Feature Descriptor Impact The Registration Accuracy in Laparoscopic Liver Surgery?. Curr Dir Biomed Eng. 2022; 8(1): 46-49.

[6] Yang Z, Simon R, Linte CA. Learning Feature Descriptors for Pre- And Intra-Operative Point Cloud Matching for Laparoscopic Liver Registration. Int J Comput Assist Radiol Surg. 2023 Jun; 18(6):1025-1032.

[7] Soler L, Hostettler A, Agnus V, Charnoz A, Fasquel J, Moreau J, et al. 3D Image Reconstruction for Comparison of Algorithm Database: A Patient Specific Anatomical and Medical Image Database. IRCAD, Strasbourg, France, Tech. Rep, 2010.

[8] Krames L, Suppa P, Nahm, W. Generation of Synthetic Data for the Comparison of Different 3D-3D Registration Approaches in Laparoscopic Surgery. Annu Int Conf IEEE Eng Med Biol Soc. 2022 Jul; 2022:1871-1874.

[9] Zhou Q, Park J, Koltun V. Open3d: A Modern Library For 3D Data Processing. 2018; arXiv preprint:1801.09847

[10] Fischler M, Bolles R. Random Sample Consensus: A Paradigm For Model Fitting With Applications to Image Analysis and Automated Cartography. Commun. of the ACM. 1981; 381-395.

[11] Gholizadeh M, Bakhshali M, Mazlooman S, Aliakbarian M, Gholizadeh F, Eslami S, et al. Minimally Invasive and Invasive Liver Surgery Based on Augmented Reality Training: A Review of the Literature. J. Robot. Surg. 2022; 17:753–763.

[12] Martínez-Cecilia D, Wicherts D, Cipriani F, Berardi G, Barkhatov L, Lainas P, et al. Impact of Resection Margins for Colorectal Liver Metastases in Laparoscopic and Open Liver Resection: A Propensity Score Analysis.Surg. Endosc. 2021; 35:809–818.

[13] Hattab G, Riediger C, Weitz J, Speidel S. A Case Study: Impact of Target Surface Mesh Size and Mesh Quality on Volume-To-Surface Registration Performance in Hepatic Soft Tissue Navigation. Int J Comput Assist Radiol Surg. 2020 Aug; 15(8):1235-1245.