Fast and accurate vision-based stereo reconstruction and motion estimation for image-guided liver surgery

Andrew D. Speers[1] ✉, Burton Ma[1], William R. Jarnagin[2], Sharifa Himidan[3,4], Amber L. Simpson[2], Richard P. Wildes[1]

[1]*Department of Electrical Engineering and Computer Science, York University, Toronto, ON, Canada*
[2]*Hepatopancreatobiliary Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA*
[3]*Department of Surgery, The Hospital for Sick Children, Toronto, ON, Canada*
[4]*Department of Surgery, University of Toronto, Toronto, ON, Canada*
✉ *E-mail: speers@eecs.yorku.ca*

Image-guided liver surgery aims to enhance the precision of resection and ablation by providing fast localisation of tumours and adjacent complex vasculature to improve oncologic outcome. This Letter presents a novel end-to-end solution for fast stereo reconstruction and motion estimation that demonstrates high accuracy with phantom and clinical data. The authors' computationally efficient coarse-to-fine (CTF) stereo approach facilitates liver imaging by accounting for low texture regions, enabling precise three-dimensional (3D) boundary recovery through the use of adaptive windows and utilising a robust 3D motion estimator to reject spurious data. To the best of their knowledge, theirs is the only adaptive CTF matching approach to reconstruction and motion estimation that registers time series of reconstructions to a single key frame for registration to a volumetric computed tomography scan. The system is evaluated empirically in controlled laboratory experiments with a liver phantom and motorised stages for precise quantitative evaluation. Additional evaluation is provided through testing with patient data during liver resection.

1. Introduction: Liver resection is the only potentially curative therapy for liver cancer but often represents a surgical challenge due to the location of tumours throughout complex, delicate vasculature [1, 2]. Image-guided surgical techniques improve intraoperative localisation [3]. While image guidance has become the standard of care in neurosurgery, the lack of non-rigid correction and reliance on static intraoperative data has hindered adoption in liver surgery. By imaging the liver in real time and providing the surgeon with continuous feedback regarding the localisation of tumours relative to adjacent major vasculature, the likelihood of an incomplete resection or inadvertent liver injury should be minimised.

A variety of approaches have been proposed to acquire real-time intraoperative data, typically in isolation. Intraoperative three-dimensional (3D) structure recovery that relies on monocular imagery has been explored but is often susceptible to error and suffers scale ambiguities without the presence of external markers [4–7]. Instead of directly computing 3D structure from monocular imagery, one approach has been explored that frames the problem of overlaying preoperative volumetric information onto the intraoperative 2D video as that of computing a projection matrix from 2D–3D correspondences between the video sequence and preoperative computed tomography (CT) [8]. Reconstruction from stereo imagery using so-called local [9, 10], semi-global [11, 12] and global approaches [13–17] also has been considered, where the main differences can be viewed as trade-offs between computational efficiency and the complexity of smoothness integration employed in image matching. The previous work has also considered a visual odometry-based approach, making use of the quadrifocal constraints to estimate binocular laparoscopic camera motion in a surgical environment [18]. Stereo imaging has emerged as a promising modality for acquiring rich continuous intraoperative surface data in neurosurgery [19, 20] but a number of challenges need to be solved for such systems to perform adequately in liver surgery. Among the most pressing and immediate challenges are that stereo reconstruction in the liver is non-trivial: the liver is a sparsely textured organ with few features to drive reconstruction

algorithms and specular reflections further compromise reconstruction accuracy.

Existing systems for optical organ registration in surgery have been proposed; however, they typically rely on techniques that may ultimately hinder their adoption in soft tissue surgery. Some deployed systems benefit from having a rigid reference frame (e.g. neurosurgery) and assume relatively little deformation of the target organ with respect to this reference frame [21, 22]. Other more recent research has considered automated rigid registration of a liver surface model with a preoperative CT scan by matching shape-based feature descriptors [23]. As rigid constraints are not always applicable to soft tissue surgery, the use of cross-modality fiducials has been proposed for use on the body [24, 25] and on the organ itself [26]. The use of fiducials creates additional invasive steps in the workflow of typical procedures to the degree that it is often desirable to avoid their use altogether. Furthermore, their utility often diminishes greatly as their placement becomes separated from the surgical surface of interest. Other systems operate through the use of manually chosen feature points on the visible anatomy [27]. The reliance on the manual selection and tracing of points in the image sequence is undesirable as it places a barrier for the autonomous operation of the system as a whole. Arguably, the most applicable approach to date is the use of intraoperative cone-beam CT (CBCT) as a bridging modality [28]. The use of CBCT allows for the use of a non-rigid biomechanically driven registration technique [29] for aligning the preoperative CT and intraoperative CBCT scans which allows for the system to compensate for the large non-rigid deformation between the two sources of data. However, the use of CBCT exposes the patient (and surgical team) to repeated doses of radiation to provide this reference.

In the light of previous research, the primary contribution of this work is a novel end-to-end system for fast surface reconstruction and motion estimation for alignment with a preoperative CT scan. Specifically, we deem this system to be 'end-to-end' as the proposed system utilises manual input for initialisation purposes only and requires neither human intervention nor does it rely on any

intraoperative bridging modalities [such as open magnetic resonance imaging (MRI), CBCT etc.] during subsequent operation. The current instantiation of the system is designed for use during intraoperative planning/exploratory phases. Extensions are underway to make the system applicable for larger portions of the surgical workflow.

Within this end-to-end system, we make the following three subcontributions. (i) To the best of our knowledge, we are the first to use an adaptive coarse-to-fine (CTF) stereo algorithm for fast and accurate 3D surface reconstruction in intraoperative imaging. The approach yields data-driven dense reconstruction by allowing coarse resolution image data to inform fine resolution reconstruction, even in low texture regions [30]. CTF processing also leads to computational efficiency, while the complementary use of adaptive windows supports the precise reconstruction of 3D boundaries. (ii) We make use of a robust, 3D motion estimator based on interframe feature matching to register a time series of reconstructions to a single key frame for registration to a volumetric CT scan. Unlike most approaches that use the iterative closest point algorithm for 3D model registration, our feature matching-based approach ensures that registration brings physically meaningful features into alignment and does so without chaining multiple incremental registrations that can rapidly lead to drift. (iii) A mask denoting the boundary of the organ of interest, the liver, is automatically maintained within the system. This has been studied in a standalone fashion (e.g. [31]) but the use of such information within the context of this system raises new insight. Maintaining a mask of the liver boundary not only allows for efficient processing of the information by restricting processing to the portion of the video stream imaging the liver but also allows for the system to take advantage of the fact that the intraoperative motion of the liver during exploratory phases is predominantly rigid, allowing for more robust motion estimation.

The system has been evaluated empirically in controlled laboratory experiments with a liver phantom placed on motorised stages for precise quantitative evaluation. Our phantom-based datasets are available to the research community and can be found at http://vision.eecs.yorku.ca/research/medical/. Additional evaluation has been undertaken with clinical data. Both evaluations take place in open liver resection conditions. Notably, while much research effort focuses strictly on laparoscopic surgeries, most liver resections are performed as open resections due to the extent and location of disease. In our centre, roughly 75% of liver resections are performed in an open environment [32]. Moreover, laparoscopic approaches are only suitable for certain cases where there is oncologic equivalency between open and laparoscopic [32]. Overall, we demonstrate a clinical stereo-based platform capable of reliably providing temporally dense 3D textured data in near real-time under realistic conditions of liver surgery.

Interestingly, the only two Food and Drug Administration (FDA)-approved liver surgery systems do not compensate for real-time motion [33]. More generally, commercial systems rely on optical tracking, which has a stated accuracy of <2 mm, as evaluated at a single point in time [34]. The errors we report are within comparable bounds and our system is capable of providing continual updates. Thus, our approach has potential to provide the precise anatomical location of tumours within the complex vasculature, in real-time, as the liver undergoes motion throughout the course of surgery.

2. Technical approach: Fig. 1 provides an overview of the system for recovering a time series of 3D surface reconstructions of a surgical scene and registering to a preoperative volumetric model. The input is a pair of synchronised images from a stereo video camera (left and right images) and a volumetric model (CT scan). The processing pipeline consists of three main components: stereo correspondence determination, 2D feature tracking and six degrees of freedom (6DOF) motion estimation. Stereo

correspondence yields a dense disparity map between points in the left and right images. The disparity map is projected into 3D space and filtered to produce a 3D surface reconstruction. 2D feature tracking is applied to the video from the right camera to provide 2D matched feature locations across the image sequence. These 2D tracks are fused with the disparity maps to produce a 3D non-rigid deformation field. 6DOF motion estimation regresses the 3D deformation field to a rigid 6DOF motion relating the current frame back to a key frame. The 3D surface reconstruction and 6DOF motion estimate are combined to place the surface into the same reference frame as a key frame. When combined with a hand registration of the key frame to the preoperative scan, the registration of incoming frames back to the key frame allows for a chained registration back to the preoperative volumetric scan of the organ of interest.

2.1. Stereo correspondence and 3D surface reconstruction: The system uses a local stereo correspondence algorithm that has been shown to provide accurate and efficient depth estimates [30]. Given a calibrated stereo pair of images, $I_l(x, y)$ and $I_r(x, y)$, the algorithm yields a disparity map, $d(x, y)$, that provides the spatial offset between corresponding points in the input pair. The disparity map is recovered by solving the optimisation problem

$$d(x, y) = \arg\max_{d_i \in D} \sum_{(u, v) \in w(x, y)} \rho\big[I_l(u, v), I_r(u + d_i, v)\big], \quad (1)$$

via search over disparities, d_i, in a specified range, D, to maximise the summed pixel-wise similarity measure, ρ, between image intensity values within a window, w, around (x, y). Although the algorithm does not rely on the choice of a specific similarity measure, ρ, our instantiation makes use of normalised cross-correlation. The algorithm employs CTF processing for efficiency, whereby initial low-resolution versions of the input images yield low-resolution disparity maps that subsequently are refined via consideration of higher-resolution images to culminate with the resolution of the original input. The algorithm also uses adaptive windows, w, that confirm to avoid smoothing across 3D boundaries. An example recovered disparity map for an input stereo pair is shown in Fig. 1.

Use of known camera calibration allows for the recovered disparities to be back-projected to a 3D point cloud. The resulting point cloud is filtered via statistical outlier removal to reject depth values that differ significantly from their neighbours [35] as well as resampled via moving least squares [36] to produce a smooth manifold surface as the final 3D surface reconstruction. A manually specified region of interest mask can be provided for an initial (key) frame in the image sequence to restrict processing to the area to be registered to the volumetric model. Following initial delineation, the mask is warped automatically to all other frames in the sequence via a robust affine estimate of image motion across the masked region [37]. An example of recovered surface reconstruction is shown in Fig. 1.

2.2. 2D feature tracking and generation of a 3D deformation field: To establish feature tracks between frames in the stereo video sequence, 2D features are tracked in one of the video streams (the right is used) and subsequently are projected to 3D. Tracking in 2D makes use of the 'good features to track' algorithm [38]. This algorithm restricts operations to feature points, (x, y), where the local image gradient structure is sufficient for stable appearance across time. Tracking is performed on the extracted features across two images, I^t and $I^{t+\delta t}$, taken at times t and $t + \delta t$, respectively, by minimising the dissimilarity measure

$$\sum_{(u, v) \in r(x, y)} [I^{t+\delta t}(u + \delta_x, v + \delta_y) - I^t(u, v)]^2, \quad (2)$$
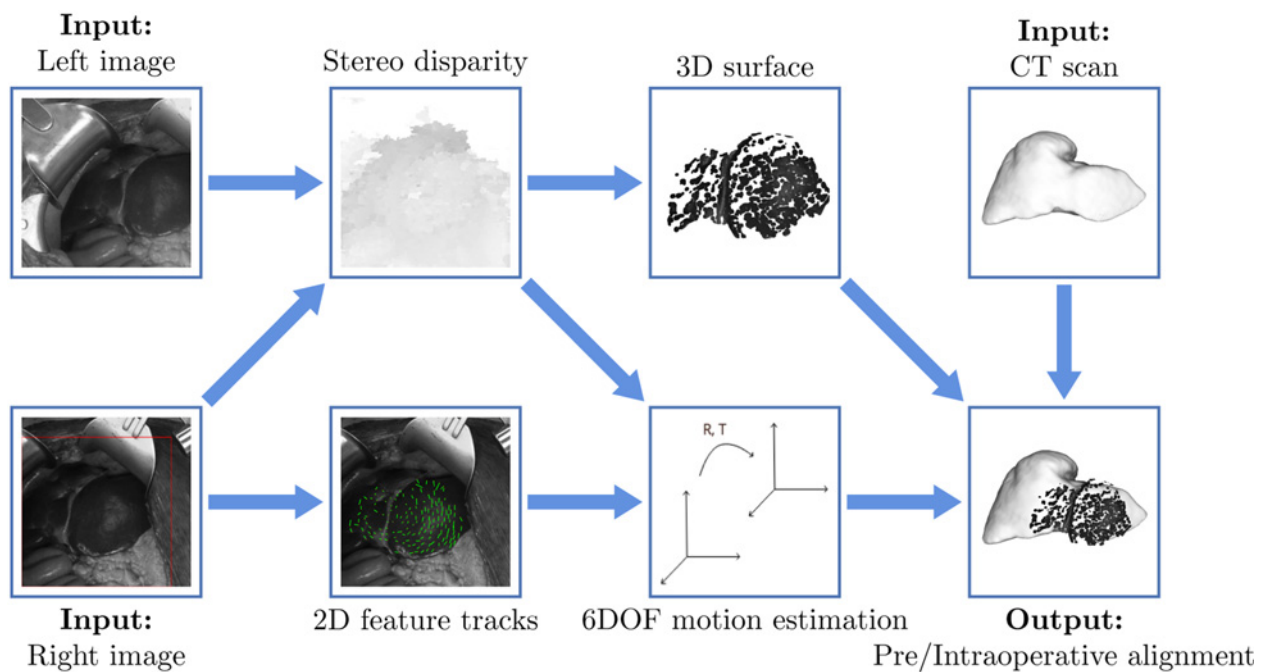
Fig. 1 *System diagram. Input images from stereo video cameras are shown in grey level with a red box indicating the region of analysis. Recovered stereo correspondences are shown as a disparity map referenced to the right image, with brighter grey levels indicating larger disparities. Green lines on tracking output show recovered 2D displacement of features across time. Pre/intraoperative alignment shows an overlay of the reconstructed surface map on the volumetric model*

over windows, r, centred at the feature points. A gradient-based solution is employed to yield the optimal feature displacement, (δ_x, δ_y), for each feature point, (x, y). While the original formulation accounted for a full affine transformation across time [38], the simpler translational formulation given here has proved to suffice for the cases of current interest. Example tracks are shown in Fig. 1.

Analogous to the back-projection of stereo disparity maps to 3D point clouds (Section 2.1), the extracted feature tracks are combined with the disparity estimates at the tracked points to yield 3D deformation fields. These fields provide a sequence of 3D feature correspondences between all frames in the sequence, capturing both the rigid and non-rigid components of the motion of the tracked region in 3D.

2.3. 6DOF motion estimation and pre/intraoperative alignment: Alignment between the time sequence of 3D surface reconstructions and the volumetric model is initialised via manual registration of a key reference frame in the reconstructions and the model. This registration is given as a 6DOF rigid transformation specified via a graphical user interface. (Future work will refine the initial alignment with an automated non-rigid registration algorithm.) Given this initial registration, all subsequent surface reconstructions from the stereo video sequence are registered to the first frame to inherit its alignment to the volumetric model.

To perform the registration across frames of the 3D surface reconstructions, a 6DOF motion between each frame and the key frame is recovered based on the previously recovered 3D deformation fields (Section 2.2). (To complement future work that considers a non-rigid initialisation to the volumetric model, the non-rigid residual to the 6DOF estimate can be considered.) The rigid transformation is recovered by a robust version of an earlier algorithm according to

$$(\boldsymbol{R}, \boldsymbol{t}) = \underset{\boldsymbol{R} \in \mathrm{SO}(3), \boldsymbol{t} \in \mathbb{R}^3}{\arg\min} \sum_{i=1}^{n} \| (\boldsymbol{R}\boldsymbol{p}_i + \boldsymbol{t}) - \boldsymbol{q}_i \|^2 \quad (3)$$

where $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$ are 3D points in two surface reconstructions that have been brought into correspondence at n locations tracked by the 3D deformation field, $\boldsymbol{R}$ is a 3×3 rotation matrix and $\boldsymbol{t}$ is a 3×1 translation vector. The rotation is found using singular value decomposition on the covariance matrix relating the two point sets after placing their centroids at the origins ($\boldsymbol{p}'$ and $\boldsymbol{q}'$). The optimal translation is then the residual created in the origin-centric point clouds after the rotation is applied (i.e. $\boldsymbol{t} = \boldsymbol{q}' - \boldsymbol{R}\boldsymbol{p}'$). The solution is made robust using the random sample consensus algorithm to minimise the effect of outlier correspondences or those most affected by any non-rigidity. Notably, registering incoming frames back to a key frame ameliorates issues of registration drift that can occur in the alternative approach of chaining sequential registrations between adjacent frames over a long sequence.

An example of final alignment between a 3D surface reconstruction and volumetric model is shown in Fig. 1. For input 640×480 images, the system executes at 8 fps on a 3.6 GHz processor. Considerable speed up is anticipated with a graphics processing unit (GPU) implementation, e.g. the slowest component of the system is stereo correspondence, which already has been ported to run on a 512 core GPU (nVIDIA GeForce GTX580) at 100 fps.

3. Empirical evaluation: The proposed system was evaluated empirically in both controlled laboratory conditions and with clinical data acquired during liver resection. Fig. 2 shows images from the test scenarios as well as acquired images, surface reconstructions and final registrations for both laboratory and operating room conditions. The laboratory dataset, depicted in Fig. 3, was acquired using a silicone liver phantom. The phantom was rigidly affixed to a motion control platform (one linear and one rotational stage – Newport Corporation, Irvine, CA) that allowed for the phantom to undergo precise motion patterns. The platform contains cross-modality features used for alignment of the initial frame to a CT scan. Each laboratory test condition consisted of a 31 frame sequences with evenly spaced samples along the motion trajectory. Three different motion profiles were tested: translation-only (1 mm increments), rotation-only (1° increments)
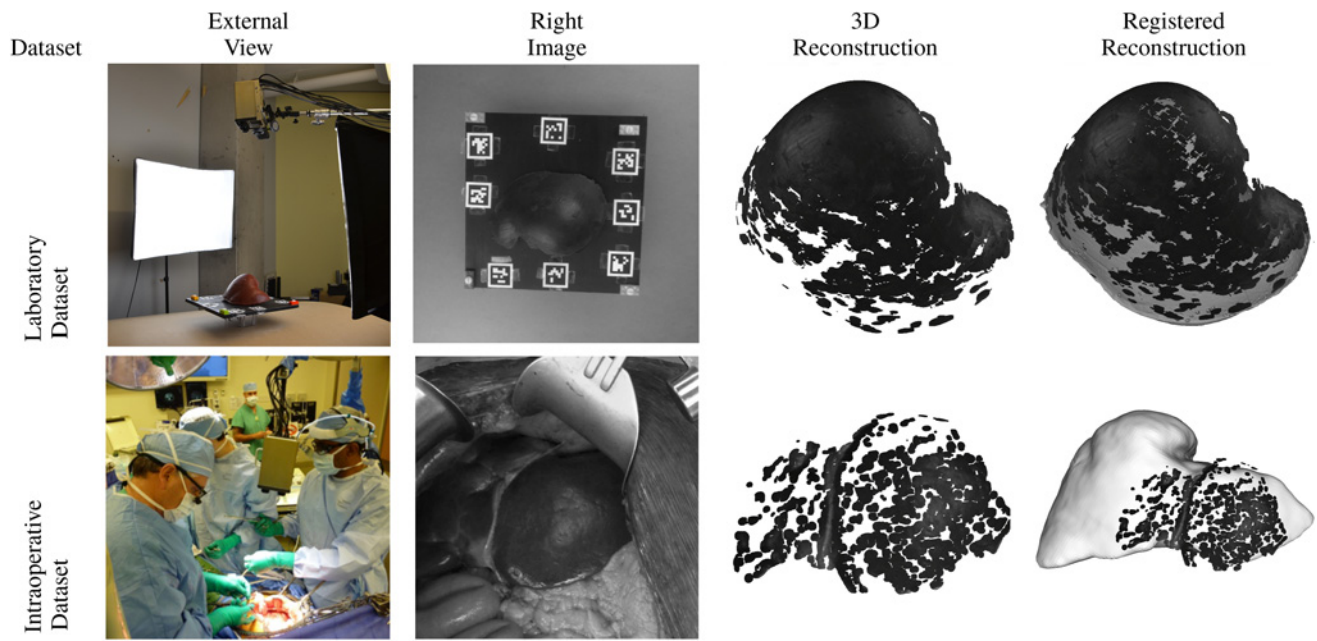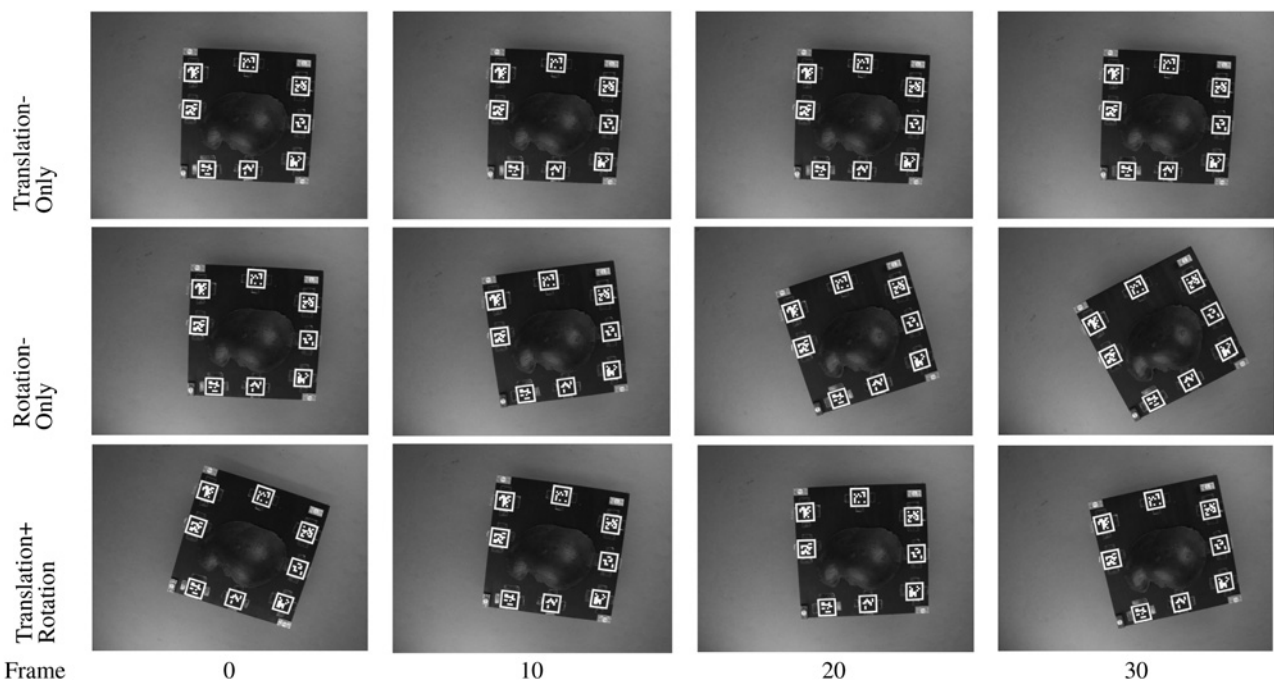
Fig. 2 *Overview of experiments*

Fig. 3 *Laboratory datasets*

and translation + rotation motions superimposed. The laboratory dataset was acquired at a standoff distance of 700 mm to the liver (similar to the target distance for the device used intraoperatively). The intraoperative data sequence was taken during an open liver resection and consisted of 100 frames. In both the laboratory and operating rooms, a calibrated stereo video camera was used for image acquisition (the gold box in the upper middle portion of the external views shown in Fig. 2).

The translation (magnitude) and rotation components of the recovered motions are reported for the laboratory datasets in Fig. 4. For each of the three tested motion profiles (rotation-only, translation-only and translation + rotation), the recovered translation

magnitude is shown in Figs. 4*a–c*, while the recovered rotation angles are shown in Figs. 4*d–f*. At any given frame, the motion was recovered with respect to the initial frame in the sequence. For rotation-only, it is seen that the recovered angle accurately tracks the true motion (interframe increments of 1°), approximately a line of slope one and the recovered translation is correctly very small. (Note that since the true translation has zero magnitude, the error between the recovered and ground truth is the same as the recovered and only the recovered is shown.) For the translation-only case, the results are exactly complementary to those of rotation-only, again showing very accurate performance. Similarly, the translation + rotation shows the desired combination
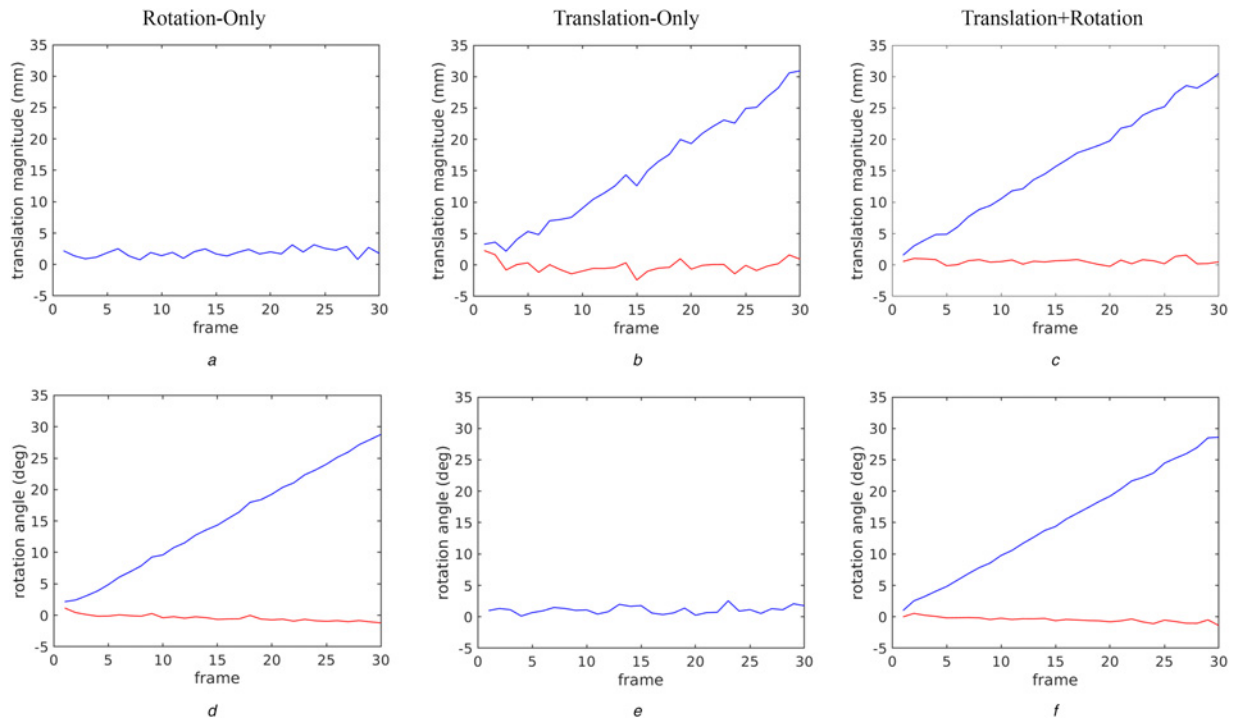
Fig. 4 *Motion recovery results. The recovered translation magnitude (a–c) and rotation (d–f) estimates are plotted for each of the three laboratory motion profiles (rotation-only, translation-only and translation + rotation). The blue line indicates the motion estimate; the red line depicts the deviation of the motion estimate from the ground truth, as actuated by the motion platform. The two plots showing a single curve correspond to measurements, where the ground truth signal was zero motion; hence, the recovered and error values are the same and only the recovered is shown*
a–c Recovered translation magnitude
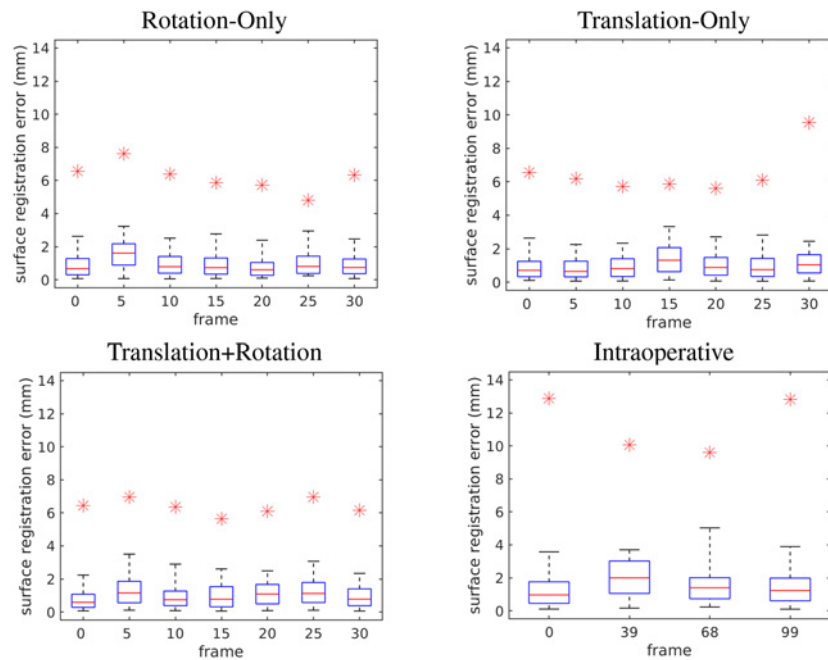*d–f Recovered rotation*

Fig. 5 *SRE (4) is shown at 5°/mm increments across the laboratory sequences. For the intraoperative data, four frames representing various amounts of respiratory motion were chosen to show the system performance over the largest portion of the organ's motion. Box plots are shown for each of the reported frames. Whiskers cover 95% of the reconstructed point clouds. The largest outlier for each dataset is indicated by a red asterisk*

of the other two cases. Notably, motion drift is not affecting the interframe estimation over these sequences owing to the motion always being computed relative to the initial frame. In contrast, typical approaches that chain transformation estimates between adjacent frames to relate a given frame back to the key frame would be susceptible to drift.

A measure of surface registration error (SRE)

$$\mathrm{SRE}_i = \|\boldsymbol{p}_i - \boldsymbol{p}_i^*\|, \tag{4}$$

where $\boldsymbol{p}_i$ is a point in the registered stereo reconstruction of the surface and $\boldsymbol{p}_i^*$ is the closest point to $\boldsymbol{p}_i$ in the preoperative

Table 1 Mean and standard deviation of reported errors

Dataset	SRE, mm		Translation magnitude, mm		Rotation angle, deg	
	Avg	Std	Avg	Std	Avg	Std
rotation-only	1.0274	0.8698	1.8975	0.6547	−0.4124	0.5126
translation-only	1.0597	0.8736	−0.1831	0.9863	1.0898	0.5744
translation + rotation	1.0785	0.8998	0.5656	0.4092	−0.4507	0.4095
intraoperative	1.6493	1.4060	n/a	n/a	n/a	n/a

volume, was calculated for all laboratory and intraoperative datasets. The plots show the distribution of SRE over all points in the reconstructed point clouds for the selected frames. Registration between CT coordinates and the initial frame of each laboratory sequence was performed in a semi-automated manner, via identification of five cross-modality features located on the platform supporting the phantom in both the stereo imagery and in the CT scan. A similar process was used for the intraoperative data after an initial hand alignment of the datasets. This procedure defines the transformation relating the initial (key) frame to the CT coordinate space. Subsequent frames were registered to the initial key frame using the approach described in Section 2.3 and then registered to CT space using the transformation relating the key frame to the CT scan. Fig. 5 shows SRE for the three laboratory datasets and the intraoperative dataset. Median errors are typically on the order of 1 mm with 95% of the points lying under 4 mm SRE across the reported datasets. Notably, the interframe registration algorithm never explicitly minimises SRE by applying shape-based registration techniques (e.g. iterative closest point); therefore, the registration is data driven with respect to the actual positions of identified features on the surface of the organ and is more likely to produce a physically meaningful registration.

Table 1 provides a summary of the average and standard deviation of the reported SRE and motion errors for the three laboratory datasets and the intraoperative dataset. Note that average errors and standard deviation of the motion estimates are not provided for the intraoperative dataset as no ground truth motion information is available.

4. Conclusion: We have presented an end-to-end system for fast and accurate 3D surface reconstruction and motion estimation for alignment with a preoperative volumetric scan. Key technical innovations include the use of an adaptive CTF algorithm for efficient and accurate 3D surface reconstruction from stereo imagery and use of a robust, feature-based 3D motion estimator for physically meaningful alignment. The system has been evaluated both quantitatively and qualitatively in controlled laboratory conditions as well as with clinical data. The results suggest the potential for integration into a clinical system. Future work will make further use of the recovered 3D deformation field (Section 2.2) to support non-rigid refinements for our 3D surface reconstructions to volumetric model alignments. An extension of this approach to include subsurface registration through the use of additional sensing modalities that support subsurface data acquisition during surgery (e.g. ultrasound, CBCT and open MRI) is also of interest. Appropriate application of intraoperative subsurface scanning and registration constraints may provide better localisation for surgeons when surgical margins are tight. Furthermore, integrating subsurface information into the surface-based registration approach is an important step for being able to measure the correlation between measurements of registration errors at the surface of the organ and the ability to target subsurface structures accurately. Finally, additional testing on clinical data is desired in order to further validate and develop these techniques.

6. Conflict of interest: None declared.

7 References

[1] Rosen C.B.: 'Management of hepatic metastases', *Cancer Control*, 1998, **5**, pp. 30–31

[2] Sardi A., Akbarov A.: 'Management of primary and metastatic tumors to the liver', *Oncology*, 1996, **10**, pp. 911–938

[3] Kingham T.P., ET AL.: '3D image guidance assisted identification of colorectal cancer liver metastases not seen on intraoperative ultrasound: results from a prospective trial', *HPB: the official J. Int. Hepato Pancreato Biliary Assoc.*, 2017, **20**, (3), pp. 260–267

[4] Hu M., Penney G., Rueckert D., ET AL.: 'Non-rigid reconstruction of the beating heart surface for minimally invasive cardiac surgery'. Medical Image Computing and Computer-Assisted Intervention, London, England, 2009, pp. 34–42

[5] Collins T., Bartoli A.: 'Towards live monocular 3D laparoscopy using shading and specularity information'. Information Processing in Computer-Assisted Interventions, Pisa, Italy, 2012, pp. 11–21

[6] Mirota D.J., Wang H., Taylor R.H.: 'A system for video-based navigation for endoscopic endonasal skull base surgery', *IEEE Trans. Med. Imaging*, 2012, **31**, (4), pp. 963–976

[7] Grasa Ó.G., Bernal E., Casado S., ET AL.: 'Visual SLAM for handheld monocular endoscope', *IEEE Trans. Med. Imaging*, 2014, **33**, (1), pp. 135–146

[8] Puerto-Souza G.A., Cadeddu J.A., Mariottini G.L.: 'Toward long-term and accurate augmented-reality for monocular endoscopic videos', *IEEE Trans. Biomed. Eng.*, 2014, **61**, (10), pp. 2609–2620

[9] Stoyanov D., Scarzanella M.V., Pratt P., ET AL.: 'Real-time stereo reconstruction in robotically assisted minimally invasive surgery'. Medical Image Computing and Computer-Assisted Intervention, Beijing, China, 2010, pp. 275–282

[10] Bernhardt S., Abi-Nahed J., Abugharbieh R.: 'Robust dense endoscopic stereo reconstruction for minimally invasive surgery'. Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 2013, pp. 254–262

[11] Vagvolgyi B., ET AL.: 'Video to CT registration for image overlay on solid organs'. AMIARCS, New York, USA, 2008, pp. 78–86

[12] Röhl S., Bodenstedt S., Suwelack S., ET AL.: 'Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration', *Med. Phys.*, 2012, **39**, (3), pp. 1632–1645

[13] Lau W.W., Ramey N.A., Corso J.J., ET AL.: 'Stereo-based endoscopic tracking of cardiac surface deformation'. Medical Image Computing and Computer-Assisted Intervention, St. Malo, France, 2004, pp. 494–501

[14] Stoyanov D., Darzi A., Yang G.Z.: 'A practical approach towards accurate dense 3D depth recovery for robotic laparoscopic surgery', *Comput. Aided Surg.*, 2010, **10**, (4), pp. 199–208

[15] Chang P.-L., Stoyanov D., Davison A.J., ET AL.: 'Real-time dense stereo reconstruction using convex optimisation with a cost-volume for image-guided robotic surgery'. Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 2013, pp. 42–49

[16] Bergeles C., Pratt P., Merrifield R., ET AL.: 'Multi-view stereo and advanced navigation for transanal endoscopic microsurgery'. Medical Image Computing and Computer-Assisted Intervention, Cambridge, United States, 2014, pp. 332–339

[17] Thompson S., Totz J., Song Y., *ET AL.*: 'Accuracy validation of an image guided laparoscopy system for liver resection'. SPIE Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling, 2015, vol. 9415

[18] Chang P.L., Handa A., Davidson A.J., *ET AL.*: 'Robust real-time visual odometry for stereo endoscopy using dense quadrifocal tracking'. Int. Conf. Information Processing in Computer-Assisted Interventions (IPCAI), Fukuoka, Japan, 2014, pp. 11–20

[19] Kumar A.N., Miga M.I., Pheiffer T.S., *ET AL.*: 'Persistent and automatic intraoperative 3D digitization of surfaces under dynamic magnifications of an operating microscope', *Med. Image Anal.*, 2015, **19**, (1), pp. 30–45

[20] Paul P., Morandi X., Jannin P.: 'A surface registration method for quantification of intraoperative brain deformations in image-guided neurosurgery', *IEEE Trans. Inf. Technol. Biomed.*, 2009, **13**, (6), pp. 976–983

[21] Kelly P.J., Kall B.A., Goerss S., *ET AL.*: 'Computer-assisted stereotaxic laser resection of intra-axial brain neoplasms', *J. Neurosurg.*, 1986, **64**, pp. 427–439

[22] Roberts D.W., Strohbehn J.W., Hatch J.F., *ET AL.*: 'A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope', *J. Neurosurg.*, 1986, **65**, pp. 545–549

[23] Robu M.R., Ramalhinho J., Thompson S., *ET AL.*: 'Global rigid registration of CT to video in laparoscopic liver surgery', *Int. J. Comput. Assist. Radiol. Surg.*, 2018, **13**, (6), pp. 947–956

[24] Ieiri S., Uemura M., Konishi K., *ET AL.*: 'Augmented reality navigation system for laparoscopic splenectomy in children based on pre-operative CT image using optical tracking device', *Pediatric Surg. Int.*, 2012, **28**, (4), pp. 341–346

[25] Tsutsumi N., Tomikawa M., Uemura M., *ET AL.*: 'Image-guided laparoscopic surgery in an open MRI operating theater', *Surg. Endosc.*, 2013, **27**, (6), pp. 2178–2184

[26] Baumhauer M., Simpfendörfer T., Muller-Stich B.P., *ET AL.*: 'Soft tissue navigation for laparoscopic partial nephrectomy', *Int. J. Comput. Assist. Radiol. Surg. (CARS)*, 2008, **3**, (3–4), pp. 307–314

[27] Su L.M., Vagvolgyi B.P., Agarwal R., *ET AL.*: 'Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3D-CT to stereoscopic video registration', *Urology*, 2009, **73**, pp. 896–900

[28] Mountney P., Fallert J., Nicolau S., *ET AL.*: 'An augmented reality framework for soft tissue surgery'. Medical Image Computing and Computer-Assisted Intervention, Cambridge, United States, 2014, pp. 423–431

[29] Oktay O., Zhang L., Mansi T., *ET AL.*: 'Biomechanically driven registration of pre- to intra-operative 3D images for laparoscopic surgery'. Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 2013, pp. 1–9

[30] Sizintsev M., Wildes R.P.: 'Coarse-to-fine stereo vision with accurate 3D boundaries', *Image Vis. Comput.*, 2010, **28**, (3), pp. 352–366

[31] Collins T., Bartoli A., Bourdel N., *ET AL.*: 'Robust, real-time, dense and deformable 3D organ tracking in laparoscopic videos'. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Athens, Greece, 2016, pp. 404–412

[32] Selby L.V., DeMatteo R.P., Tholey R.M., *ET AL.*: 'Evolving application of minimally invasive cancer operations at a tertiary cancer center', *J. Surg. Oncol.*, 2017, **115**, (4), pp. 365–370

[33] Simpson A.L., Kingham T.P.: 'Current evidence in image-guided liver surgery', *J. Gastrointest. Surg.*, 2016, **20**, pp. 1265–1269

[34] Northern Digital Inc.: Polaris Series Optical Tracking System. https://www.ndigital.com/msci/products/polaris-series, accessed 31 July 2017

[35] Rusu R.B., Marton Z.C., Blodow N., *ET AL.*: 'Towards 3D point cloud based object maps for household environments', *Robot. Auton. Syst.*, 2008, **56**, (11), pp. 927–941

[36] Alexa M., Behr J., Cohen-Or D., *ET AL.*: 'Computing and rendering point set surfaces', *IEEE Trans. Vis. Comput. Graph.*, 2003, **9**, (1), pp. 3–15

[37] Black M.J., Anandan P.: 'The robust estimation of multiple motions: parametric and piecewise-smooth flow fields', *Comput. Vis. Image Underst.*, 1996, **63**, (1), pp. 75–104

[38] Shi J., Tomasi C.: 'Good features to track'. Computer Vision and Pattern Recognition, Seattle, Washington, United States, 1994, pp. 593–600