

Benchmarking Video With The Surgical Image Registration Generator (SIRGn) Baseline

Michael Barrow¹, Nelson Ho², Alric Althoff³, Peter Tueller¹, and Ryan Kastner¹

¹ University of California, San Diego, La Jolla CA 92093, USA
{mbarrow,ptueller,rkastner}@eng.ucsd.edu

² VMware, Palo Alto CA 94304, USA
hon@vmware.com

³ Leidos, San Diego CA 92127, USA
aalthoff@eng.ucsd.edu

Abstract. Augmented Reality (AR) surgical image guidance overlays preoperative data into the surgeon’s view in real time during the procedure. Non-rigid 3D registration is a critical and often challenging step for AR surgical image guidance. Since surgical environments vary greatly and registration must be done quickly and accurately, it is unlikely that one registration technique will work well over different surgical scenarios. Unfortunately, it is currently challenging to evaluate the accuracy and effectiveness of 3D registration techniques on surgical scenes. In this work, we provide a novel method to benchmark quality of non-rigid 3D surface registration. Our method provides a triangular mesh overlay representing the quality of registration and can highlight areas of unacceptably poor registration performance given some specified tolerance. We use the method to evaluate the quality of two existing non-rigid registration approaches on surgical video.

Keywords: Benchmarking, Registration, Augmented Reality, Image Guidance

1 Introduction

Augmented Reality (AR) has significant potential for enhancing surgical procedures. For example, the ability to superimpose landmarks like major internal vessels onto the surface of an organ provides immense value for surgical image guidance. There are many therapeutic uses for AR guidance in laparoscopic surgery [1–5] and other closely related intervention types such as telerobotic surgery where AR enhances the pre-existing feedback from robotic tools [6, 7].

A key challenge for AR surgery is registration, which provides the relationship between a preoperative model (e.g., CT or MRI scan) and the surgical video feed. Fig. 1 illustrates the registration process. The three images/frames on the left are collected in real-time during the procedure from a camera. Each frame must be registered onto the preoperative 3D model on the right. After that, features from the preoperative 3D model are overlaid onto the surgeons view.

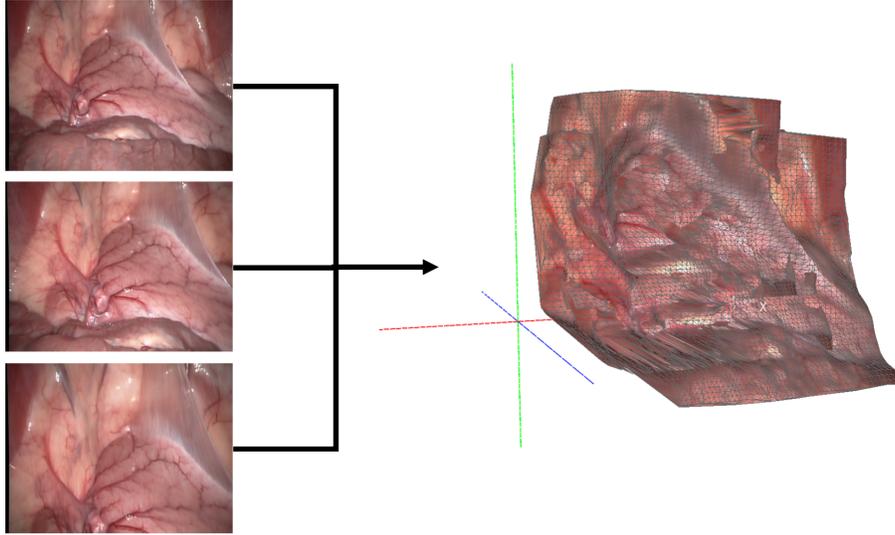


Fig. 1: Registration maps video frames (left) onto a 3D model (right). The video frames should be mapped in real-time onto the 3D model allowing any features from the model to be projected into the surgeons view.

Surgical augmented reality is particularly challenging since it must be done in real time with high accuracy. A surgical scene can be extremely complex: tissues deform, lighting varies dramatically, and features quickly come in and out of the field of view. Every laparoscope, every lighting solution, and every type of surgery have unique idiosyncrasies. Consequently, custom registration algorithms are needed for the unique challenges arising from a given combination of factors in a particular surgery. It can be difficult to tune the performance of a registration algorithm to the surgeon’s needs since benchmarks that account for specific idiosyncrasies of that surgery do not exist. Because of these challenges, surgery-specific heuristics are needed to perform high quality dense surface registration in real-time. However, it is impossible to measure the effectiveness of these heuristics without a baseline quality measure. A useful quality measure must be *dense*, *granular*, and *convenient*.

Many quality metrics have been applied to clinical registration. The canonical approach is Target Registration Error (TRE) which involves computing $L2$ or some other distance between corresponding feature points after registration [8–12]. Feature separation distances can be highly representative of registration error if the features are appropriately chosen, but such metrics have no concept of a global coordinate system to track drift in a stream of registrations. Besides distance measures, global similarity metrics based on intensity have broad clinical applications [13–16]. These qualitative measures combine intensities of

the registered images into one image. A clinician can then judge the quality of registration through experienced observation of the resulting overlay of images. Because of this reliance on clinical experience, intensity metrics are not used in applications where registration accuracy is critical as they cannot represent the warp function in a physically meaningful way.

Landmark registration metrics judge registration quality by an objective function weighted by pre-defined feature points. These consider rigid registration [17] and/or fiducial markers [18] rather than tissue only landmarks and can be used to focus the metric on registration regions of clinical interest. Hoffmann et al. [19] develop such a quality metric. Unfortunately, they do not attempt large surface area coverage, they rely on a small number of landmarks, and their technique is applied to CT scans rather than video. Recently Thompson et al. [20] proposed an AR specific registration evaluation focused on accurately positioning sub-surface landmarks. Unfortunately shifting focus from surface registration to sub-surface features requires that assumptions are made on the biomechanics of the organ when estimating landmark positions. Because such assumptions often break down [21], quantitative surface registration quality metrics are more robust.

Surgical Image Registration Generator (SIRGn) is a novel dense, granular baseline quality metric for video registration algorithms. Fig. 2 shows how SIRGn works. It evaluates the registration of a set of images/frame (Fig 2a) by creating a heat map that shows regions of poor registration quality (Fig. 2b) that can be subsequently overlaid onto the scene (Fig. 2c) for visual inspection of the quality of result of the registration algorithm.

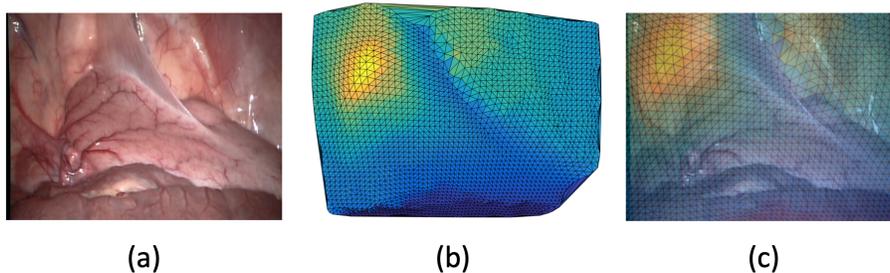


Fig. 2: Illustration of the SIRGn concept. Assume a 3D registration algorithm is run on the video frame shown in Fig. 2a. SIRGn evaluates the quality of the registration by creating a covariance mesh (Fig. 2b) and overlaying that as a “heat map” (Fig. 2c) where high heat indicates poor registration quality, e.g., as seen in the upper left corner.

The novelty of SIRGn lies in the fact that it combines the physically meaningful metric of TRE with the clinically useful dense image coverage of an inten-

sity type measure. Additionally, unlike the aforementioned approaches, SIRGn is well suited for image guidance benchmarking since real time registration algorithms are evaluated using a global coordinate system, which allows it to compare across multiple frames. Our method can be applied in situations where no suitable database of videos exists or where synthetic simulation is intractable. Also, it can be applied to many types of surgery and it does not rely on physical simulation or biomechanical tissue information.

In this manuscript we describe the SIRGn method and an experiment that uses SIRGn to compare performance of two non-rigid 3D registration algorithms on laparoscopic video from a database [22]. Our experiment models the exploratory phase of AR image guidance where several approaches are tried to accurately track the clinically significant portion of a surgical scene over time. Specifically we show that SIRGn is advantageous in evaluating the accuracy of different registration methods by comparing a SIRGn work flow with the conventional TRE and RMSE error metrics. Registration performance is evaluated using landmark correspondences [23]. Although it is desirable in clinical image registration to use expert landmark correspondences, our goal is convenience⁴ and so we provide tools for automated global reference feature extraction from 2D and 3D laparoscopic video.

The remainder of this paper is organized as follows. Section 2 gives an overview of the baseline generation method and evaluation of 3D registration quality using SIRGn as a baseline. Section 3 gives an appraisal of SIRGn as a simple and flexible baseline generation method. We conclude in section 4.

2 Method

SIRGn provides a way to evaluate a given registration algorithm on a surgical 3D video data set. For this, we must create a metric that describes the quality of the registration. An ideal metric would work for any registration algorithm and data set. It would allow us to compare the quality of different registration algorithms on that data set. And it would provide a location specific quality measure. We describe our metric in the context of stereo laparoscopic cameras since we anticipate SIRGn’s greatest applicability to be in image guided surgery. However our methodology can be applied to conventional 2D surgical video data.

The insight motivating SIRGn is that surfaces in surgical scenes are smooth with very few exceptions. Therefore given a smooth surface, a point that is heavily warped by registration with respect to surrounding “known-good” points is less likely to be accurate. In order to formalize this intuition, we rely on two sets of points. The first, G , is a set of expert landmarks that we assume as correct a priori; these could be human labeled by a surgeon or provided by a reliable feature detection algorithm. The second, C , contains points that require heavy warping to align well on a global model M . Points in C are chosen by searching points enclosed within the triangulation of G . Specifically, a set of candidate

⁴ Example code and data available at: “<https://github.com/KastnerRG/SIRGn.git>”

points to be registered is masked using a triangle from G to project onto M space. After registration the point that required the largest registration warping transform to M space is added to C . Any point that requires substantial warping is likely due to an incorrect registration, and similarly any points that require a small amount of warping are likely correctly registered. Thus, the key is to determine how to evaluate these warp functions for any registration algorithm and to derive a method to efficiently locate points C that must be heavily warped. We describe this in more detail in the following.

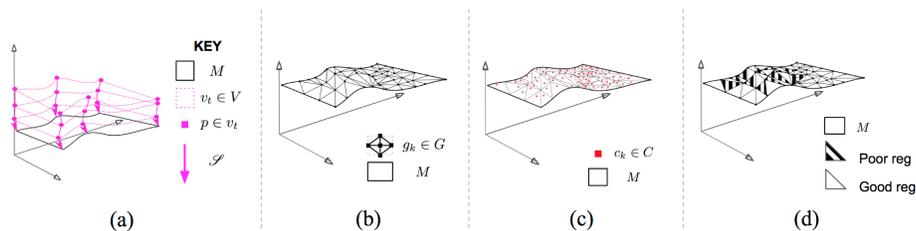


Fig. 3: Steps to compute the SIRGn quality metric are, in order: (a) Construct a 3D model M and warp \mathcal{S} using points $g_k \in G$ and all 3D scans $v_t \in V$. (b) Project expert-labeled landmarks $g_i \in G$ onto M . (c) Find heavily-warped registered points $c_i \in C$. (d) Label a mesh with vertices G showing registration quality. The labels/coloring are a function of G , C and ε tolerance and can be visualized, e.g., as done in figs. 5d to 5f.

The four major components to SIRGn are shown graphically in Fig. 3. Fig. 3a computes a non-rigid warping function \mathcal{S} that projects G from the video frames V to the 3D model M . That is, the warp function \mathcal{S} describes the relationship between the landmarks G in the stereo video frames V and those same landmarks on the 3D model M . One can view M as containing the average location of all the landmarks within G over time. Fig. 3b computes a Delaunay triangulation between each landmark g_i on the 3D model M . Fig. 3c finds the most heavily warped points for each surface triangle. Fig. 3d assigns a value to each triangle computed from these sets to faces of a Delaunay triangulation to form a “correlation mesh”. This is the output that indicates how well the worst case points correlate with the landmarks. If there is high correlation, then the worst case points correspond well to the assumed high quality landmarks, and thus are more likely to indicate that the registration was accurate. This correlation mesh is very useful for comparing the performance of different registrations with one another since it highlights regions of relatively poor performance.

Once a candidate algorithm has been selected using the correlation mesh, SIRGn can be re-run and set to report mean registration error of the G triangulation for a candidate registration in another mesh. We set an acceptable error threshold based on the registration accuracy desired. This is typically defined by the needs of a clinical AR application, i.e., 5mm is typical abdominal surg-

eries [24]. Unacceptable regions of the candidate registration are highlighted in a second mesh. This work flow is demonstrated in Section 3.

2.1 SIRGn Registration Quality Measure

In order to use SIRGn, an evaluator must first run a global registration algorithm, which we call ALG. ALG is chosen to be highly accurate, i.e., real time constraints do not apply in order to maximize the registration quality. Because AR registration is performed in \mathcal{R}^3 , SIRGn should be run after a stereo reconstruction algorithm on stereoscopic laparoscopes although it can be run directly on 3D scopes. ALG returns a dense global registration model M by performing a global registration optimization that minimizes the transform distance of the set of G points for all 3D video frames $v_t \in V$, such that

$$M = \text{ALG}(V, G) \quad (1)$$

Global Warping Function: Once we have computed M , we obtain a dense benchmark registration over M . To do this, we compute a function \mathcal{S} mapping from scans V to the space of M to allow every point in 3D depth scans $v_t \in V$ to be mapped to the dense surface. We choose the method used by Global Non-Rigid Alignment (GNRA) [25] using thin plate spline (TPS) interpolation that is regularized to produce a smooth function. We prefer this method as a warp baseline both for its stability and the guarantee that \mathcal{S} is a smooth bijective transform. The resulting \mathcal{S} is the global warping function, such that $\mathcal{S}(p; t) = \bar{p} \in M$ for any point p in a 3D scan v_t . \mathcal{S} is therefore useful when building the C set as detailed later.

Landmark distance vectors: SIRGn builds a set \mathcal{D}_G of landmark distance vectors consisting of one vector ℓ_k for each expert landmark in the set of all landmarks G . Note that landmark points g_k in some 3D scan v_t correspond to each other throughout scans $v_t \in V$, and to one \bar{g}_k in the space of the global model M . We compute index t of these vectors ℓ_k as the distance between $g_k \in v_t$ and $\bar{g}_k \in M$. That is,

$$\ell_k[t] = \|g_k - \bar{g}_k\|_2, \text{ s.t. } g_k \in v_t \quad (2)$$

for each global landmark $\bar{g}_k \in M$. Then $\mathcal{D}_G = \{\ell_k : 1 \leq k \leq |G|\}$ is the set of landmark distance vectors.

Warping distance vectors: SIRGn computes a mesh of M via Delaunay triangulation using G as mesh vertices. Then, for each triangular face Δ_{c_k} of the mesh we define points $c_{k,t}$ in frames v_t as the *most heavily warped in* Δ_{c_k} at time t . Formally,

$$c_{k,t} = \operatorname{argmax}_{p \in \Delta_{c_k}} \mathcal{S}^{-1}(p; t) \quad (3)$$

Note that since \mathcal{S} is bijective and everywhere differentiable, \mathcal{S}^{-1} is well-defined for each t , and Eqn. (3) can be solved using standard optimization techniques. We choose these points $c_{k,t}$ because a point on the model surrounded by expert landmarks on the surface of an organ that undergoes a different warp than these

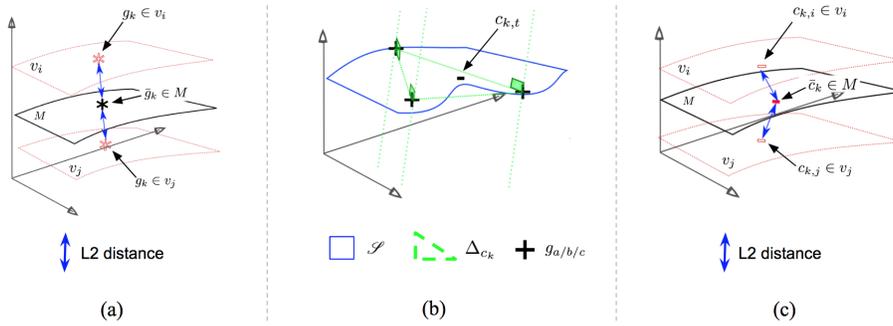


Fig. 4: (a) Indices i and j of the landmark distance vector ℓ_k corresponding to \bar{g}_k are computed as the Euclidean (L2) distance from the corresponding points $g_k \in v_i$ and $g_k \in v_j$. (b) The most heavily warped point $c_{k,t}$ are computed for a single time point t as the maximum warp \mathcal{S} required to send the point $p \in \Delta_{c_k}$ to the model M . (c) Warping distance vectors ω_k are computed similarly to landmark distance vectors, except $c_{k,t}$ is used instead.

landmarks is unlikely to track the tissue surface well. We form a set C_t of these “worst points” $c_{k,t}$, one for each triangle in the mesh, by projecting c_k ’s enclosing triangle to form a planar region of interest (ROI) and then performing gradient ascent on \mathcal{S}^{-1} within these ROIs. Once C_t is determined, the set \mathcal{D}_{C_t} of warping distances is computed identically to \mathcal{D}_G with $c_{k,t}$ and $\bar{c}_{k,t}$ replacing g_k and \bar{g}_k in Eqn. (2).

Correlation mesh: In order to make our metric dense over the registration surface, we define Delaunay mesh face-values as the linear correlation between vectors $\omega_{k,t} \in \mathcal{D}_{C_t}$, and averages of those vectors corresponding to the vertices of Δ_{c_k} from \mathcal{D}_G . That is,

$$\text{SIRGn}(\Delta_{c_k}) = \frac{1}{n} \sum_t \text{corr}(\omega_{k,t}, \hat{\ell}_k) \quad (4)$$

where $\text{corr}(u, v)$ is the linear correlation (normalized covariance) between vectors u and v , and $\hat{\ell}_k = (\ell_a + \ell_b + \ell_c)/3$ where $\ell_a, \ell_b, \ell_c \in \mathcal{D}_G$ are the three vectors corresponding to those g_k that are the vertices of Δ_{c_k} , and n is the number of frames containing c_k . Good registration in the local region Δ_{c_k} is indicated by a large correlation between the vectors for that particular Δ_{c_k} . In this way we obtain a quality score that is both *dense*, because the whole model surface is covered, and *granular*, so that each region can be specifically highlighted for easy localization of unacceptable registration errors.

Labeling: This correlation mesh can be transformed into a binary accept/reject by taking a tolerance parameter ε to define what level of error is acceptable. We can also display the linear correlation on the mesh faces directly via coloring for final visualization. Labeling the Delaunay triangle set of G is the final step in generating the SIRGn quality metric.

3 Experiments and Results

We track respiration motion in a laparoscopic surgery scene accurately in order to best place image guidance landmarks on a live patient. We use SIRGn and the conventional metrics of RMSE and TRE to compare two 3D surface registration algorithms on laparoscopic video from the Hamlyn data set [22]. Our goal is to use each of the comparison metrics to understand which registration algorithm was best suited to tracking the non-rigid respiration motion in the scene. Our findings detailed in this section showed that SIRGn meshes offered a simple way to compare registration in specific regions of interest in a case where traditional metrics were more difficult to interpret.

We ran all metrics on two algorithms under test. The first, "*ICP(V)*", is a non-rigid ICP variant based on Amberg's work [26]. The second, "*GEO(V)*", uses Chen's geodesic approach [27]. Geodesic registration is suited to non-rigid problems but is typically computed offline due to its complexity. However, a real time implementation was recently described, making it interesting to compare with ICP [28].

SIRGn was run on the algorithms under test using the following steps. First the $ALG(V, G)$ inputs were generated from a laparoscopic video data base [22] using a stereo reconstruction specialized to laparoscopic video (V) [29] and a robust global SIFT tracker (G). Although mixed manual methods for extracting ground truth are more desirable, these must be applied during video acquisition which was not possible in our case [30, 31]. Next, C and M were generated by running $ALG(V, G)$.

Correlation: We applied Eqn. 4 to the output from *ICP(V)* and *GEO(V)* to generate relative quality meshes as shown in Figs. 5b and 5c. Yellow is weak correlation with the benchmark and deep blue is strong correlation with the benchmark. The color map of the figures was normalized to the algorithms under test for ease of visual comparison.

Labeling: After using SIRGn to compare *ICP(V)* and *GEO(V)* quality meshes, we ran SIRGn with a labeling threshold ϵ . The ϵ was set to reject error $\geq 10, 5$ and 3mm as shown in figs. 5d and 5e and fig. 5f respectively. Mesh triangle registration errors of ϵ and above were greyed out.

SIRGn runtime varied from 5 to 12 minutes depending on $ALG(V, G)$ input size. An example implementation is available under the BSD license at *Omitted for blind review*. Figure 4 summarizes the results of all metrics where $|G| = 1155$ and $|V| = 180$ which were the largest input we used for 6 seconds of video.

Traditional Metrics: Traditional metrics were computed frame to frame as they are not designed for global registration. Completely global RMSE and TRE metrics (fig. 5i) cannot represent time and it was not possible to use them in assessing how stable either *ICP(V)* or *GEO(V)* might have been. In addition registration drift could not be accounted for which led to misleadingly low error reports. Even when a time dimension is added to RMSE as in fig. 5g it was not possible to isolate the respiratory ROI.

Although we could use our G and C points to define mean TRE per frame as shown in fig. 5h, it was still not trivial to isolate the ROI. In contrast, the SIRGn

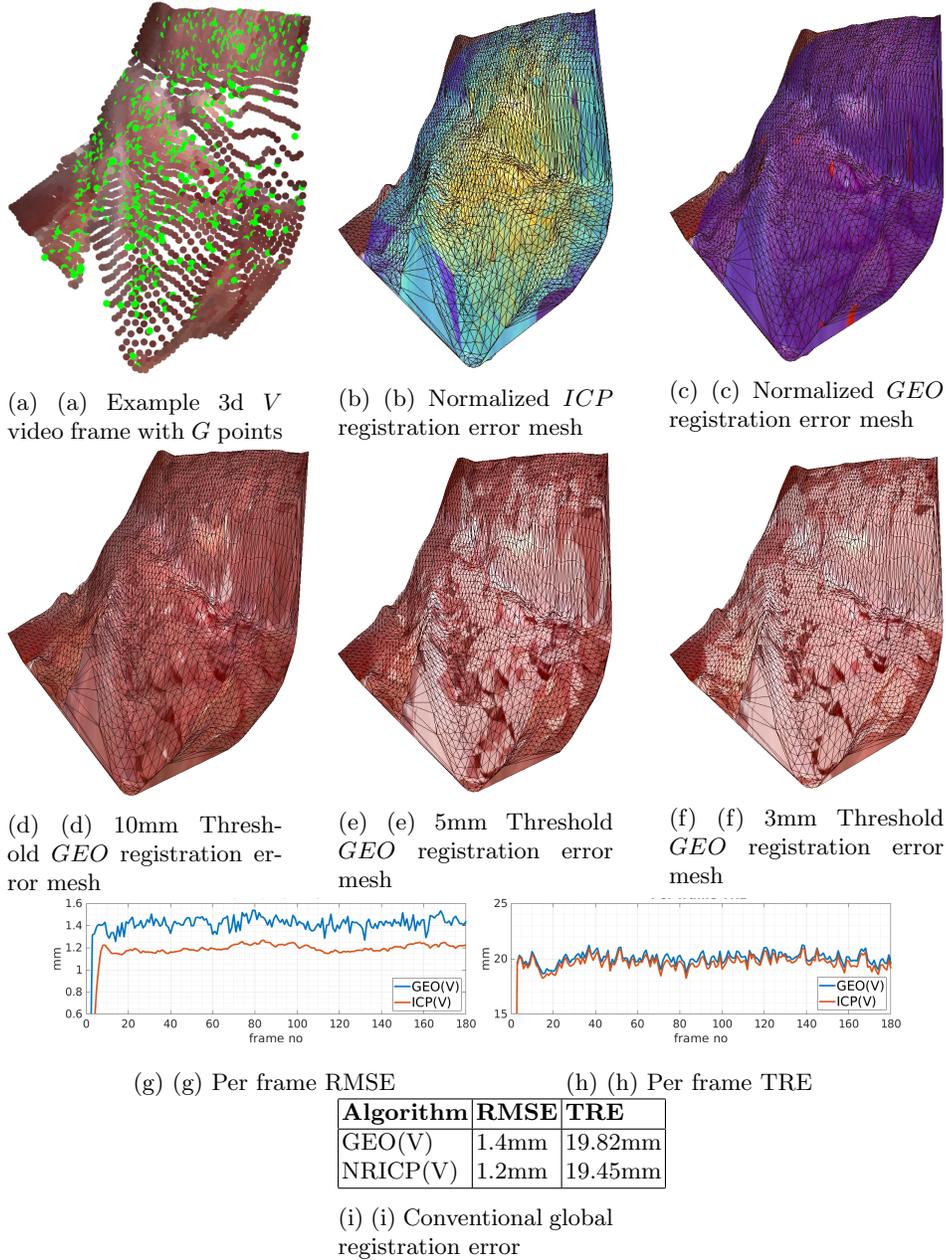


Fig. 4: Example use of SIRGn to compare two non-rigid 3D registration approaches (figs. 5a to 5c), alongside traditional error metrics (figs. 5g to 5i). Figure 5a shows a priori known good points (G) rendered on one video frame in green. Figure 5b and fig. 5c are the quality meshes for the $ICP(V)$ and $GEO(V)$ video registrations respectively. According to SIRGn, $ICP(V)$ showed a far greater variation than $GEO(V)$ from the baseline in the region of video with most movement. Figure 5d, fig. 5e and fig. 5d show the results of SIRGn set to color the $GEO(V)$ error mesh white. based on different triangle registration error thresholds. Important parts of the scene can be inspected at the desired threshold setting to see if registration is acceptable. Comparatively, traditional metrics shown in fig. 5i, fig. 5g and fig. 5h do not show the variation of registration performance with area which was less informative than SIRGn for evaluating quality of registration around the region of interest in the video clip.

overlays in figs. 5b and 5c clearly show regional variations of registration quality between $ICP(V)$ and $GEO(V)$ on the respiratory region, which was colored yellow for $ICP(V)$. SIRGn was most reflective of the more rigid registration $ICP(V)$ implemented with respect to $GEO(V)$ of all the evaluated metrics. In addition the ε threshold of SIRGn allowed us to see that the GEO algorithm was likely unsuitable for an abdominal AR guidance application requiring sub 5mm accuracy. This is in contrast to our evaluations with traditional metrics of RMSE and TRE where it is difficult to know if sub 5mm accuracy is possible in regions of interest.

4 Conclusion

We describe SIRGn – a novel metric for evaluating the quality of nonrigid registration algorithms on surgical 3D video. SIRGn indicates the quality of the registration for different parts of the 3D model by generating an overlay mesh that highlights areas where the registration approach succeeds and fails. This can serve as a way to evaluate different registration algorithms (both new and old) in various surgical scenarios. The ultimate goal is to provide a metric by which different registration algorithms crucial to enabling surgical augmented reality can be compared and evaluated. Our experiments demonstrate how SIRGn can be used to compare different registration methods. SIRGn evaluates a given registration algorithm by creating a mesh of registration quality over a 3D surgical video sequence using known good landmarks and our novel warp variance interpolation. We demonstrate how a SIRGn mesh can show relative registration quality between algorithms under evaluation, and also how an error tolerance can be specified when generating a registration quality mesh to highlight regions of unacceptable registration accuracy for a given application.

References

1. Pessaux, P., Diana, M., Soler, L., Piardi, T., Mutter, D., Marescaux, J.: Towards cybernetic surgery: robotic and augmented reality-assisted liver segmentectomy. *Langenbeck's archives of surgery* **400**(3) (2015) 381–385
2. Ieiri, S., Uemura, M., Konishi, K., Souzaki, R., Nagao, Y., Tsutsumi, N., Akahoshi, T., Ohuchida, K., Ohdaira, T., Tomikawa, M., et al.: Augmented reality navigation system for laparoscopic splenectomy in children based on preoperative ct image using optical tracking device. *Pediatric surgery international* **28**(4) (2012) 341–346
3. Onda, S., Okamoto, T., Kanehira, M., Fujioka, S., Suzuki, N., Hattori, A., Yanaga, K.: Short rigid scope and stereo-scope designed specifically for open abdominal navigation surgery: clinical application for hepatobiliary and pancreatic surgery. *Journal of hepato-biliary-pancreatic sciences* **20**(4) (2013) 448–453
4. Kowalczyk, J., Meyer, A., Carlson, J., Psota, E.T., Buettner, S., Pérez, L.C., Farritor, S.M., Olynykov, D.: Real-time three-dimensional soft tissue reconstruction for laparoscopic surgery. *Surgical endoscopy* **26**(12) (2012) 3413–3417

5. Schoob, A., Kundrat, D., Kleingrothe, L., Kahrs, L.A., Andreff, N., Ortmaier, T.: Tissue surface information for intraoperative incision planning and focus adjustment in laser surgery. *International journal of computer assisted radiology and surgery* **10**(2) (2015) 171–181
6. Hughes-Hallett, A., Mayer, E.K., Pratt, P., Mottrie, A., Darzi, A., Vale, J.: The current and future use of imaging in urological robotic surgery: a survey of the european association of robotic urological surgeons. *The International Journal of Medical Robotics and Computer Assisted Surgery* **11**(1) (2015) 8–14
7. Volonté, F., Pugin, F., Buchs, N.C., Spaltenstein, J., Hagen, M., Ratib, O., Morel, P.: Console-integrated stereoscopic osirix 3d volume-rendered images for da vinci colorectal robotic surgery. *Surgical innovation* **20**(2) (2013) 158–163
8. Fitzpatrick, J.M.: *Detecting failure, assessing success*. CRC Press New York (2001)
9. Hong, J., Matsumoto, N., Ouchida, R., Komune, S., Hashizume, M.: Medical navigation system for otologic surgery based on hybrid registration and virtual intraoperative computed tomography. *IEEE Transactions on Biomedical Engineering* **56**(2) (2009) 426–432
10. Shekhar, R., Dandekar, O., Bhat, V., Philip, M., Lei, P., Godinez, C., Sutton, E., George, I., Kavic, S., Mezrich, R., et al.: Live augmented reality: a new visualization method for laparoscopic surgery using continuous volumetric computed tomography. *Surgical endoscopy* **24**(8) (2010) 1976–1985
11. Krucker, J.F., LeCarpentier, G.L., Fowlkes, J.B., Carson, P.L.: Rapid elastic image registration for 3-d ultrasound. *IEEE Transactions on Medical Imaging* **21**(11) (2002) 1384–1394
12. Grachev, I.D., Berdichevsky, D., Rauch, S.L., Heckers, S., Kennedy, D.N., Caviness, V.S., Alpert, N.M.: A method for assessing the accuracy of intersubject registration of the human brain using anatomic landmarks. *NeuroImage* **9**(2) (1999) 250–268
13. Nanayakkara, N.D., Chiu, B., Samani, A., Spence, J.D., Samarabandu, J., Fenster, A.: A twisting and bending model-based nonrigid image registration technique for 3-d ultrasound carotid images. *IEEE transactions on medical imaging* **27**(10) (2008) 1378–1388
14. Tamura, Y., Sugano, N., Sasama, T., Sato, Y., Tamura, S., Yonenobu, K., Yoshikawa, H., Ochi, T.: Surface-based registration accuracy of ct-based image-guided spine surgery. *European spine journal* **14**(3) (2005) 291–297
15. Parsai, E.I., Ayyangar, K.M., Dobelbower, R.R., Siegel, J.A.: Clinical fusion of three-dimensional images using bremsstrahlung spect and ct. *The Journal of Nuclear Medicine* **38**(2) (1997) 319
16. Nelson, S.J., Day, M.R., Buffone, P.J., Wald, L.L., Budinger, T.F., Hawkins, R., Dillon, W.P., Huhn, S., Prados, M.D., Chang, S., et al.: Alignment of volume mr images and high resolution [18f] fluorodeoxyglucose pet images for the evaluation of patients with brain tumors. *Journal of computer assisted tomography* **21**(2) (1997) 183–191
17. Fitzpatrick, J.M., West, J.B., Maurer, C.R.: Predicting error in rigid-body point-based registration. *IEEE transactions on medical imaging* **17**(5) (1998) 694–702
18. West, J.B., Fitzpatrick, J.M., Toms, S.A., Maurer Jr, C.R., Maciunas, R.J.: Fiducial point placement and the accuracy of point-based, rigid body registration. *Neurosurgery* **48**(4) (2001) 810–817
19. Hoffmann, C., Krause, S., Stoiber, E.M., Mohr, A., Rieken, S., Schramm, O., Debus, J., Sterzing, F., Bendl, R., Giske, K.: Accuracy quantification of a deformable image registration tool applied in a clinical setting. *Journal of applied clinical medical physics* **15**(1) (2014) 237–245

20. Thompson, S., Schneider, C., Bosi, M., Gurusamy, K., Ourselin, S., Davidson, B., Hawkes, D., Clarkson, M.J.: In vivo estimation of target registration errors during augmented reality laparoscopic surgery. *International journal of computer assisted radiology and surgery* **13**(6) (2018) 865–874
21. Wong, V.W.S., Vergniol, J., Wong, G.L.H., Foucher, J., Chan, H.L.Y., Le Bail, B., Choi, P.C.L., Kowo, M., Chan, A.W.H., Merrouche, W., et al.: Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. *Hepatology* **51**(2) (2010) 454–462
22. London, I.: Hamlyn centre laparoscopic/endoscopic video datasets (2019). URL <http://hamlyn.doc.ic.ac.uk/vision/>. [Accessed 15 Jan 2019]
23. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine & Biology* **54**(7) (2009) 1849
24. Bernhardt, S., Nicolau, S.A., Soler, L., Doignon, C.: The status of augmented reality in laparoscopic surgery as of 2016. *Medical image analysis* **37** (2017) 66–90
25. Brown, B.J., Rusinkiewicz, S.: Global non-rigid alignment of 3-d scans. In: *ACM Transactions on Graphics (TOG)*. Volume 26., ACM (2007) 21
26. Amberg, B., Romdhani, S., Vetter, T.: Optimal step nonrigid icp algorithms for surface registration. In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE* (2007) 1–8
27. Chen, Q., Koltun, V.: Robust nonrigid registration by convex optimization. In: *Proceedings of the IEEE International Conference on Computer Vision*. (2015) 2039–2047
28. Barrow, M., Burns, S.M., Kastner, R.: A fpga accelerator for real-time 3d non-rigid registration using tree reweighted message passing and dynamic markov random field generation. In: *2018 28th International Conference on Field Programmable Logic and Applications (FPL), IEEE* (2018) 335–3357
29. 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. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer* (2013) 42–49
30. Padoy, N., Blum, T., Ahmadi, S.A., Feussner, H., Berger, M.O., Navab, N.: Statistical modeling and recognition of surgical workflow. *Medical image analysis* **16**(3) (2012) 632–641
31. Bouarfa, L., Jonker, P.P., Dankelman, J.: Discovery of high-level tasks in the operating room. *Journal of biomedical informatics* **44**(3) (2011) 455–462