Brain-Shift Correction with Image-Based Registration and Landmark Accuracy Evaluation

Wolfgang Wein(✉)

ImFusion GmbH, Munich, Germany
wein@imfusion.de

Abstract. We describe an algorithm and its implementation details for automatic image-based registration of intra-operative ultrasound to MRI for brain-shift correction during neurosurgery. It is evaluated on a public database of 22 surgeries for retrospective evaluation, with a particular focus on choosing the appropriate transformation model and designing the most meaningful evaluation strategy. The method succeeds in a fully automatic fashion in all cases, with an average landmark registration error for the rigid model of 1.75 mm.

1 Introduction

For brain tumor resection, navigated surgery is an established approach, allowing for pre-operative MRI images and planned structures to be available during surgery, registered to various intra-operative tools by means of an external tracking system. However, the accuracy of such systems is often affected by changing soft tissue throughout the course of surgery. Intra-operative 3D freehand ultrasound, tracked with the same localizer than the other surgical tools, allows for updated information about the resection site during all stages of surgery. Unfortunately, only surgeons who are also expert users of medical ultrasound embrace this approach currently. For it to be more widely used and applicable, more automatic handling of the intra-operative ultrasound is crucial. In particular, automatic image-based registration of the 3D ultrasound data to pre-operative MRI is a first important prerequisite. While a number of such algorithms have been presented in the past, an automatic thorough evaluation remains challenging. In that context, a new public database of ultrasound and MRI volumes acquired during glioma surgery [1] was made available, which includes multiple sets of anatomical landmark points to allow for Ground Truth matching. In the following, we present the results of our image-based registration algorithm on this data. The method is based on the multi-modal similarity metric denoted linear correlation of linear combinations, or LC^2 [2] and has recently been used in a first live evaluation during surgery [3]. Other popular approaches for image-based registration utilize gradient orientations [4], or self-similarity [5].

D. Stoyanov et al. (Eds.): POCUS 2018/BIVPCS 2018/CuRIOUS 2018/CPM 2018,
LNCS 11042, pp. 146–151, 2018.
https://doi.org/10.1007/978-3-030-01045-4_17

2 Method

In summary, the registration algorithm is based on optimizing a specific ultrasound-tailored multi-modal similarity metric denoted LC^2 on 3D patches of the MRI and ultrasound volumes. A non-linear optimization algorithm changes the values of a parametric transformation model to maximize it. While the LC^2 formulation is invariant with respect to modality-specific differences in appearance, it should be restricted to volume areas whose structures match when registered. Therefore it is advisable to remove the bright stripe on the skin surface of the ultrasound volumes, since it is missing in MRI. This would be trivial to do before ultrasound volume compounding; since the available data only has reconstructed volumes, a slightly more complex approach is necessary. The ultrasound volumes can have any orientation with respect to how the voxels are arranged in memory. Therefore we traverse the volumes along all axis both forward and reverse, and sum over occurrences of black, followed by intensities above a threshold for at least 1.5 mm. From the six directions, we choose the one with the most such occurrences and delete those skin sections with a 4 mm thickness. This cuts sufficient surface if the angle to the volume axis is oblique, and slightly more than necessary if it aligns with it.

The registration is implemented in the proprietary ImFusion SDK with full OpenGL-based GPU acceleration. The ultrasound volume is assigned as fixed volume, resampled to 0.5 mm (half the MRI voxel size), and properly zero-masked. The chosen similarity metric patch-size is 7^3 voxels, as optimized in prior work. Two non-linear optimizers successively operate on the parameters of a rigid pose from the initialization as provided by the navigation system. The first is a global DIRECT (DIviding RECTangles) sub-division method [6] searching on translation only, followed by a local BOBYQA (Bound Optimization BY Quadratic Approximation) algorithm [7] on all six parameters. Optionally, the local optimizer then executes another search on full affine parameters in order to accomodate non-uniform scaling and shearing of the data, or optimizes any other parametric transformation model. As an alternative to that, a dense-deformable Demons algorithm can create local forces based on the LC^2 patch values, iteratively updating and smoothing those forces until convergence. The average computation time is ≈ 20 s on a laptop with a NVIDIA GTX 1050 mobile GPU for the global and local rigid models, with an extra 10 s for the Demons algorithm if used. A dedicated workstation GPU is typically around 3–4 times faster, allowing for almost instant results during surgery after the ultrasound volumes have been acquired.

3 Evaluation

Table 1 shows our results, put alongside the best rigid and affine transformation that can be derived from the Ground Truth landmarks. In most of the cases, the error after rigid image-based registration is within a millimeter of the smallest rigid transformation that can be fit to the landmarks. The remaining ones

Table 1. Registration results in terms of residual landmark errors in mm.

Case	Landmarks			Image-based		
	Before	Rigid	Affine	Rigid	Affine	Demons
1	1.82	1.21	1.07	1.72	1.55	2.03
2	5.68	1.37	1.10	2.53	2.54	2.54
3	9.58	0.88	0.76	1.33	1.37	1.22
4	2.99	1.11	0.98	1.65	2.27	1.72
5	12.02	1.16	0.93	1.50	1.76	1.92
6	3.27	1.14	0.81	1.67	1.65	1.66
7	1.82	1.37	1.21	1.57	2.29	1.63
8	2.63	1.37	1.06	1.94	1.92	1.99
12	19.68	1.01	0.92	1.06	1.09	1.07
13	4.57	1.03	0.95	3.74	3.07	3.02
14	3.03	1.03	1.00	1.20	1.12	1.20
15	3.21	1.48	1.28	1.91	1.83	1.89
16	3.39	1.09	0.90	1.24	1.30	1.34
17	6.39	1.30	1.02	1.71	1.35	1.64
18	3.56	0.85	0.76	1.24	1.42	1.59
19	3.28	0.97	0.81	2.12	2.69	2.85
21	4.55	0.95	0.74	1.87	1.67	1.84
23	7.01	0.99	0.70	1.89	1.47	1.85
24	1.10	0.83	0.74	1.12	1.01	1.10
25	10.06	1.32	0.87	2.78	2.55	2.12
26	2.83	1.18	0.98	1.36	1.24	1.48
27	5.76	1.18	1.05	1.44	1.70	2.22
Mean	**5.37**	**1.13**	**0.94**	**1.75**	**1.77**	**1.81**

were visually inspected and all but one deemed well registered too; only case 13 exhibits an apparent slight rotation with respect to the Ground Truth (while still improving on the original landmark error before registration).

A more thorough accuracy evaluation is complicated by the fact that the landmarks themselves only have a limited accuracy. The residual errors denoted in the left columns of Table 1 are contributed to by two factors, namely (1) the localization error of experts selecting the point correspondences, and (2) the misfit of the rigid and affine transformation models on the given data. The average landmark error of $\approx$1 mm would hence be an upper bound for the localization error given a perfect rigid or affine registration. Most likely though, such a transformation model is not entirely sufficent even on the pre-resection data at hand due to tissue deformations, and possibly tracking and calibration errors.

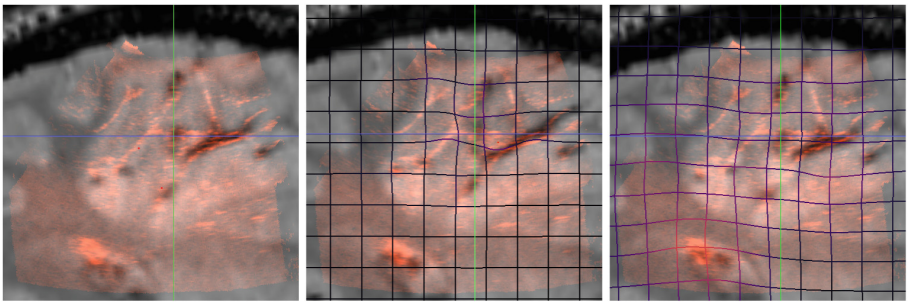
To investigate further, we are using the hypothesis that both the landmark errors as well as the LC^2 similarity metric reveal the accuracy of the registration down to a certain scale. If they both improve (in a reciprocal way of course, since we are comparing error values to a similarity metric), the chosen transformation model most likely improved the alignment. Table 2 shows the LC^2 similarity measure values for the same transformation models as shown in the error table, with one addition: We use a parametric deformation similar to radial basis functions (RBF), where an inverse distance formulation is used to interpolate between the landmark locations in the most smooth possible way. Hence, the landmark error here is forced to zero up to a numerical epsilon due to the distance field inversion. However, as can be seen in the RBF column in the table, the LC^2 value for this model is slightly worse than the affine fit. This suggests that here, the deformation might be overfit to the point correspondences

Table 2. LC^2 similarity metric on all data for the different transformation models.

Case	Landmarks				Image-based		
	Before	Rigid	Affine	RBF	Rigid	Affine	Demons
1	0.157	0.174	0.179	0.174	0.183	0.192	0.196
2	0.145	0.179	0.174	0.194	0.195	0.195	0.215
3	0.167	0.239	0.246	0.233	0.245	0.268	0.276
4	0.169	0.188	0.191	0.194	0.196	0.230	0.222
5	0.182	0.248	0.255	0.246	0.265	0.297	0.317
6	0.164	0.184	0.187	0.184	0.200	0.216	0.213
7	0.138	0.144	0.150	0.145	0.147	0.166	0.168
8	0.203	0.231	0.239	0.233	0.247	0.273	0.291
12	0.149	0.269	0.276	0.266	0.271	0.296	0.309
13	0.174	0.190	0.203	0.192	0.227	0.229	0.245
14	0.162	0.230	0.235	0.227	0.240	0.242	0.258
15	0.163	0.184	0.189	0.187	0.194	0.194	0.221
16	0.191	0.245	0.245	0.253	0.253	0.271	0.285
17	0.177	0.247	0.255	0.253	0.258	0.282	0.289
18	0.186	0.229	0.221	0.235	0.258	0.273	0.279
19	0.195	0.232	0.230	0.227	0.240	0.265	0.270
21	0.149	0.175	0.178	0.173	0.201	0.207	0.225
23	0.172	0.240	0.252	0.260	0.252	0.264	0.278
24	0.192	0.204	0.208	0.202	0.209	0.215	0.205
25	0.136	0.173	0.199	0.177	0.170	0.179	0.187
26	0.217	0.276	0.298	0.278	0.287	0.308	0.312
27	0.153	0.198	0.199	0.203	0.203	0.242	0.227
Mean	**0.170**	**0.213**	**0.218**	**0.215**	**0.225**	**0.241**	**0.250**

including their localization errors as opposed to actual structural deformations. It is also illustrated by the example in Fig. 1(b). Apart from that, the similarity consistently increases from the initial transformation, over rigid, to affine. Likewise, it increases for the image-based rigid and affine model with the highest improvement on the dense demons model. Most notably, the similarity improvements are each significantly higher than the corresponding change in landmark errors in Table 1, again suggesting a better alignment towards the right columns.

In terms of a statistical analysis, in Table 1 the only columns that are not significantly different are the various image-based transformation models (three right-most columns, Rigid/Affine/Demons), i.e. the landmark errors are all in the same order of magnitude. In Table 2, columns 2–4 (landmarks rigid vs. RBF), 3–4 (landmarks affine vs. RBF) and 3–5 (landmarks affine vs. image-based rigid) are insignificantly different. A paired Wilcoxon test with a p-value threshold of 0.01 was used. All other results are significantly different, hence it is obvious that registering landmarks only versus using our image-based method produces different results, each favoring their metric.



(a) (b) (c)

Fig. 1. Registration on case 1, **(a)** rigid with our method, **(b)** RBF on landmarks, **(c)** Demons with our method.

4 Conclusion

We have presented an algorithm for fully automatic registration of pre-resection ultrasound to pre-operative MRI volumes during brain surgery, which improves the registration for all 22 cases. For the majority of cases, the used transformation models yield landmark errors only slightly worse than the best landmark fit (in the order of magnitude of the localization errors). Since it is also known that the fiducial registration errors (here on the provided landmarks) cannot be used to reliably predict the error of a clinical target such as the tumor center [8], further validation should be performed on Ground Truth segmentations of the actual tumor mass, e.g. by computing Dice overlap values.

It will certainly also be quite interesting to locally compare regions where the LC^2 similarity metric causes a systematic shift away from the true alignment, because the local linearity assumptions are violated by the complex underlying imaging physics. Such occurrences could be improved upon with Deep Learning techniques, either by (1) learning a weighting of how reliable the similarity metric is, or (2) directly learning a more non-linear version of the similarity metric that is also taking more image content than its own patch into account.

Regarding the transformation model, it is apparent that the pre-resection data is mostly but not entirely rigid; here, a custom parametric model could be developed, for example using parameter reduction techniques on all available landmarks, also taking the position of the ultrasound probe into account where compression is strongest. To eventually achieve higher clinical impact, techniques should be developed to also continuously register ultrasound volumes during and after resection and always visualize information from the pre-operative plan accordingly. Besides a more elaborate transformation model, this will also require to exclude the resection site from registration, since it only has been changed in one of the images.

Acknowledgments. Thanks to the entire team at ImFusion, both for some additional software development supporting this work, as well as providing valuable feedback.

References

1. Xiao, Y., Fortin, M., Unsgård, G., Rivaz, H., Reinertsen, I.: Retrospective evaluation of cerebral tumors (RESECT): a clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries. Med. Phys. **44**, 3875–3882 (2017)
2. Wein, W., Ladikos, A., Fuerst, B., Shah, A., Sharma, K., Navab, N.: Global registration of ultrasound to MRI using the LC^2 metric for enabling neurosurgical guidance. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8149, pp. 34–41. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40811-3_5
3. Reinertsen, I., Iversen, D., Lindseth, F., Wein, W., Unsgård, G.: Intra-operative ultrasound based correction of brain-shift. In: Intraoperative Imaging Society Conference, Hanover, Germany (2017)
4. De Nigris, D., Collins, L., Arbel, T.: Fast rigid registration of pre-operative magnetic resonance images to intra-operative ultrasound for neurosurgery based on high confidence gradient orientations. Int. J. Comput. Assist. Radiol. Surg. **8**, 649–661 (2013)
5. Heinrich, M., et al.: MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. **16**, 1423–1435 (2012)
6. Jones, D., Perttunen, C., Stuckmann, B.: Lipschitzian optimization without the lipschitz constant. J. Optim. Theory Appl. **79**, 157 (1993)
7. Powell, M.J.: The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives. Cambridge Report NA2009/06, University of Cambridge (2009)
8. Fitzpatrick, J.: Fiducial registration error and target registration error are uncorrelated. In: Miga, M., Wong, I., Kenneth, H. (eds.) Proceedings of the SPIE, vol. 7261 (2009)