Evaluation of Brain MRI Alignment with the Robust Hausdorff Distance Measures[*]

Andriy Fedorov[1], Eric Billet[1], Marcel Prastawa[2], Guido Gerig[2],
Alireza Radmanesh[3], Simon K. Warfield[4], Ron Kikinis[3],
and Nikos Chrisochoides[1]

[1] Center for Real-Time Computing, College of William and Mary, USA
[2] Scientific Computing and Imaging Institute, University of Utah, USA
[3] Surgical Planning Laboratory, Harvard Medical School, USA
[4] Computational Radiology Laboratory, Harvard Medical School, USA

Abstract. We present a novel automated method for assessment of image alignment, applied to non-rigid registration of brain Magnetic Resonance Imaging data (MRI) for image-guided neurosurgery. We propose a number of robust modifications to the Hausdorff distance (HD) metric, and apply it to the edges recovered from the brain MRI to evaluate the accuracy of image alignment. The evaluation results on synthetic images, simulated tumor growth MRI and real neurosurgery data with expert-identified anatomical landmarks, confirm that the accuracy of alignment error estimation is improved compared to the conventional HD. The proposed approach can be used to increase confidence in the registration results, assist in registration parameter selection, and provide local estimates and visual assessment of the registration error.

1   Introduction

The objective of this work is the development of a novel metric for evaluating the results of pairwise mono-modal Non-Rigid Image Registration (NRR). An important feature of the proposed metric is the quantitative measure of the misalignment between the two images, with the goal to estimate the registration error. The specific application where we consider such metric to be of particular importance is the non-rigid registration of brain Magnetic Resonance Imaging (MRI) data during image-guided neurosurgery. This work is motivated by the difficulty of selecting the optimum parameters for NRR and the lack of "ground truth" that can be used for intra-operative evaluation of the registration results during the course of the neurosurgery.

Image registration for image-guided neurosurgery aims the alignment of the high-quality pre-operative MRI data with the scans acquired intra-operatively (lower-quality images), for subsequent visualization of the registered data to assist with the tumor targeting during the resection. Non-rigid registration [1] is essential for this application because the brain shift cannot be recovered accurately using rigid or affine transformations [2]. A number of methods for non-rigid

---

registration of brain MRI have been developed [3,4]. One of the challenges in NRR is the evaluation and validation of the registration results, e.g., see Christensen et al. [5]. Before a registration method is used in clinical studies, it can be validated on the cases, where the "ground truth" is available (e.g., phantoms, animal studies, cadavers, manually pre-labeled datasets) [1]. However, validation of the results obtained *during* the patient studies is quite limited.

In the case of non-rigid registration of brain MRI, a widely accepted approach to validation requires an expert (ideally, a group of independent experts) to identify landmark points in the pairs of images before and after the registration, or outline the corresponding anatomical regions. The accuracy can then be assessed by the overlap of the corresponding regions and landmarks. Although this approach provides possibly the best precision, the accuracy estimation is available only at the landmark locations. The procedure is also very time-consuming, and is not practical when the results need to be validated within the short period of time (e.g., during the neurosurgery, when the time allowed for NRR is limited to 5-10 minutes), or when it is not feasible to have expert involved (e.g., when the results from a large number of different registration methods need to be compared, or while performing large-scale parametric studies).

Because of the difficulties in performing intra-operative validation for patient studies, the results of NRR can be assessed by evaluating certain formalized metrics. Christensen et al. [5] summarize a number of such *success criteria*. However, not all of these criteria can be applied intra-operatively (e.g., the relative overlap metric requires segmentation of both images). Those criteria that can be evaluated intra-operatively (i.e., transformation transitivity and consistency metrics) cannot be used to estimate local alignment error. Finally, although the failure of NRR can sometimes be detected, there is no *sufficient* metric to conclude success of the NRR. This provides motivation for the development and study of new metrics for the NRR accuracy assessment.

In this paper we consider the use of an image similarity metric, different from the one minimized during NRR, as the measure of image alignment to assess the NRR accuracy. The concept of performing NRR assessment in such a way was previously suggested in [2,6]. One of the common deficiencies of many image similarity metrics (e.g., Normalized Cross-Correlation, or Mutual Information) is that their value does not quantify the alignment error in terms of Euclidean distance. Thus, given the selected similarity metric is robust and reliable, one can track the improvement in image alignment, but cannot speculate about the degree of misalignment. Our approach to address this deficiency uses similarity metrics that are derived from the definition of the Hausdorff Distance (HD) [7].

The Hausdorff distance is a very common measure in pattern recognition and computer vision to measure mismatch between the two sets of points. A number of methods have been proposed to identify features (points, edges, lines) in medical images [8,9]. The resulting feature images can be used as input for the HD measure. The HD is not based on point correspondence, which makes it somewhat tolerant to the differences in the two sets of features compared. However, it is highly sensitive to noise. A large number of robust modifications

to the HD have been proposed to suppress the noise and improve robustness of the HD, since one of the first papers in this direction by Huttenlocher et al. [7]. The most recent surveys of the modifications to the HD are available in [10,11]. The value of the HD is derived from Euclidean distances between the two point sets. Nevertheless, the HD has found limited use as a measure for the evaluation of image alignment. Peng et al. [12] used a robust version of the HD to register outlines of brain in two dimensions. Morain-Nicolier et al. [13] applied the HD to quantify brain tumor evolution. Finally, Archip et al. [2,6] assess the performance of non-rigid image registration of brain MRI with the 95% partial HD, but do not discuss the reliability of this approach.

To the best of our knowledge, the HD-based approach to image alignment assessment has not been comprehensively evaluated before. A number of the robust versions of the HD exist, but they have not been evaluated for 3d images and in the context of NRR for medical imaging. In this paper we evaluate the recent advances in the development of robust HD, and apply these techniques to evaluate the accuracy of pairwise alignment for brain MRI subject to non-rigid deformation. Based on the results of our evaluation, the presented approach significantly improves the accuracy of the previously used alignment evaluation metrics based on the conventional HD. The implementation of the presented approach is available as open source software, accompanied by the detailed description of the parameters we used to obtain the reported results on publicly available BrainWeb data [14].

2   Methods

The methods developed in this paper focus primarily on registration assessment for image guided neurosurgery. The objective of the NRR is to align the pre-operative image with the intra-operative data. Consequently, the objective of the evaluation procedure is to confirm that alignment indeed improved following NRR, and quantify the level of mis-alignment before and after registration. The first image, which is called *fixed image*, is acquired intra-operatively, and shows the brain deformation. For the purposes of assessment, the second image can either be the *floating image* (pre-operative image rigidly aligned with the fixed image), or the *registered image* (the result of registering the floating and target images). By evaluating the alignment of fixed vs. floating and fixed vs. registered images we can assess the error of alignment before and after registration, respectively. However, the formulation of the problem remains the same.

Given two images, $\mathcal{I}$ and $\mathcal{J}$, the objective is to derive the point-wise alignment error. Let $\mathcal{A}$ and $\mathcal{B}$ be the binary images with the feature points extracted from $\mathcal{I}$ and $\mathcal{J}$ respectively, and $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$ be the sets of points that correspond to the non-zero voxels in $\mathcal{A}$ and $\mathcal{B}$, respectively. Next we consider the sequence of the HD-based similarity measures with increasing robustness.

The directed HD between the two sets of points $h(A, B)$ is defined as the maximum distance from any of the points in the first set to the second set, and the HD between the two sets, denoted $H(A, B)$, is the maximum of these distances [7]:

$$h(A, B) = \max_{a \in A}(d(a, B)), \text{ where } d(a, B) = \min_{b \in B}\|a - b\|,$$

$$H(A, B) = \max(h(A, B), h(B, A)).$$

In the case of perfect correspondence between the points in the sets A and B (i.e., point a in the set A corresponds to the same image feature as point b in the set B), $H(A, B)$ would be the maximum (global), alignment error between the two images. This is the first problem in using the HD for alignment assessment, as it can only estimate the maximum error. The second problem obviously comes from the sensitivity of the metric to noise and lack of point correspondence: the estimated value of the error will not correspond to the maximum error in the general case. Simple versions of the robust HD measure were proposed to alleviate this problem. The partial Hausdorff distance is defined as a quantile of the ranked distances between the two point sets, originally proposed by Huttenlocher et al. [7]. Archip et al. [2,6] use 95%-HD, which is defined as the 0.95-quantile partial distance between the two sets. However, 95%-HD is a global measure, and does not allow to assess the error locally without modifications to the calculation procedure.

The local-distance map (LDMap) proposed by Baudrier et al. [11] for the of 2d images extends the definition of the HD, and allows to derive the *local* measure of dissimilarity between the two binary images:
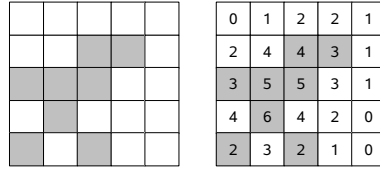
$$\forall x \in \mathbf{R}^3 : H_{loc}(x) = |\mathbf{1}_{\mathcal{A}(x)} - \mathbf{1}_{\mathcal{B}(x)}| \times \max(d(x, A), d(x, B)), \tag{1}$$

where $\mathcal{A}(x)$ is the voxel value at location x, and $\mathbf{1}_{\mathcal{A}(x)}$ is a function which has value 1 if $\mathcal{A}(x)$ is non-zero, and 0 otherwise. H_{loc} is symmetric, and it is connected to the conventional HD definition by the relation $H(A, B) = \max(H_{loc}(A, B))$ [11].

The advantage of H_{loc} (LDMap) is that it can be used for localized estimation of the alignment error. Ideally, the value of H_{loc} should be the same as the distance between the corresponding points in the images. However, because there is no point correspondence used in the HD definition, the values of H_{loc} would significantly deviate from the values of the alignment error, in the general case.

We attempt to add the notion of point correspondence to the definition of the LDMap by using greyscale modification of the HD originally proposed by Zhao et al. [10] for matching 2d images corrupted by noise. We transform the input binary images, produced by the feature detection procedure, into greyscale $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$. These greyscale images have the same size as the initial binary images, with each voxel initialized to the number of non-zero voxels in the neighborhood of the corresponding binary image pixel. A 2d example of greyscale image construction is shown in Figure 1. Let $\tilde{B}$ be the set of points that correspond to non-zero pixels in $\tilde{\mathcal{B}}$. The directed distance $d(a_g, \tilde{B})$, where g is the greyscale value at voxel a in the greyscale image computed from $\mathcal{A}$, is now defined as the distance from a to the closest voxel in $\tilde{B}$ with the greyscale value within the tolerance t from the value of g in $\tilde{\mathcal{B}}$ (we used $t = 2$):

$$d(a_g, \tilde{B}) = \min_{b_{g'} \in \tilde{B}} \|a_g - b_{g'}\|, \ g - t \leq g' \leq g + t.$$

0	1	2	2	1
2	4	4	3	1
3	5	5	3	1
4	6	4	2	0
2	3	2	1	0

Fig. 1. Left: Binary image. Right: Corresponding greyscale image.

The greyscale local HD can now be computed based on this updated point distance definition:

$$\forall x \in \mathbf{R}^3 : GH_{loc}(x) = |\mathbf{1}_{\mathcal{A}(x)} - \mathbf{1}_{\mathcal{B}(x)}| \times \max(d(x_g, \tilde{A}), d(x_g, \tilde{B})). \qquad (2)$$

We define the greyscale Hausdorff distance (GHD) $GH(A, B)$ between the two binary images as $GH(A, B) = \max(GH_{loc}(A, B))$.

Additional processing of the greyscale HD can increase its robustness further. We define robust greyscale HD locally based on the least trimmed squares robust statistics [15] on the values of GH_{loc} in the region around each feature point. The robust greyscale HD $RGH_{loc}(x)$ is calculated as the average of the ordered values of $GH_{loc}(x)$ in the fixed size window centered at x, after discarding 20% percent of the top distance values within this window (trimmed mean value). Similarly to the HD and GHD, we define the robust greyscale Hausdorff distance (RGHD) $RGH(A, B) = \max(RGH_{loc}(A, B))$.

Prior to edge detection, we smooth the input images using edge-preserving anisotropic diffusion (variance 1.0, conductance 0.5, time-step 0.0625), followed by adaptive contrast equalization [16]. Without such preprocessing the edges detected in the images can have very small overlap even with the perfect alignment. Edge detection is done with the Canny edge detector [8]. We use adaptive selection of the edge detection thresholds based on the binary search to have similar number of edges in both images. Insight Toolkit (ITK) [16] is used for all image processing operations. The reader is referred to [14] for the details of parameter selection and the open source implementation of the presented technique.

3   Results

We evaluate the presented methods using three benchmarks: (1) synthetic non-rigid deformation; (2) synthetic tumor growth; (3) real data from image-guided neurosurgery with the expert-placed anatomical landmarks. In each case, the performance of an evaluation metric is measured as its ability to recover the deformation magnitude (thus, misalignment error value) at the locations of the image where such deformation is known (all points for benchmarks (1) and (2), and selected landmark points for (3)). The questions to answer are how the values of H_{loc}, GH_{loc} and RGH_{loc} locally compare to the ground truth alignment
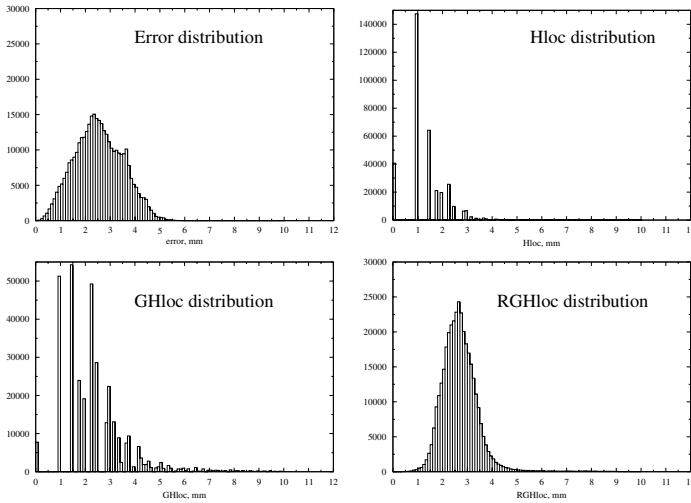
Fig. 2. Distribution of the error and the HD, GHD, and RGHD values for the same synthetic deformation case (BrainWeb, sum of Gaussian deformations, variance 5)

error, and how the robust versions of the HD (GHD and RGHD) compare with the conventional HD.

Synthetic nonrigid deformation. BrainWeb MRI simulator[1] was used to create two normal subject T1 images (0% and 9% noise) with 1 mm slice thickness, and 0% intensity non-uniformity. We applied synthetic deformation to the image, using the framework described in [17,14]. The deformation at points sampled on a regular grid is calculated as a sum of Gaussian kernels, and deformation at non-grid image locations are interpolated with thin-plate splines [16]. The magnitude of deformation can be controlled by changing the variance of Gaussian kernel. Local alignment accuracy was estimated between the undeformed image with 0% noise and deformed images with 0% and 9% noise. The accuracy of alignment at an image location (error) is the magnitude of the synthetic deformation vector at that location.

The distributions of the true error values, and H_{loc}, GH_{loc}, and RGH_{loc} are shown in Figure 2. Evidently, RGH_{loc} is a significantly more accurate approximation of the error distribution. Robustness can also be compared by looking at the percentage of outliers within local distance estimations, shown in Figure 3. We define outlier as a local estimation that exceeds the true error value at a point by more than 2 mm (deformation field is in physical space, while the HD-estimation is in 1-mm voxel space, thus errors as large as $\sqrt{3}$ cannot be prevented). With increasingly larger deformations, the ratio of outliers is also increasing. The contribution of the outliers in the conventional HD is increasing rapidly for larger deformations. The robust metrics have by far less outliers, which is reflected in the more stable behavior of GHD and RGHD in comparison to the HD: RGHD

---

[1] http://www.bic.mni.mcgill.ca/brainweb/
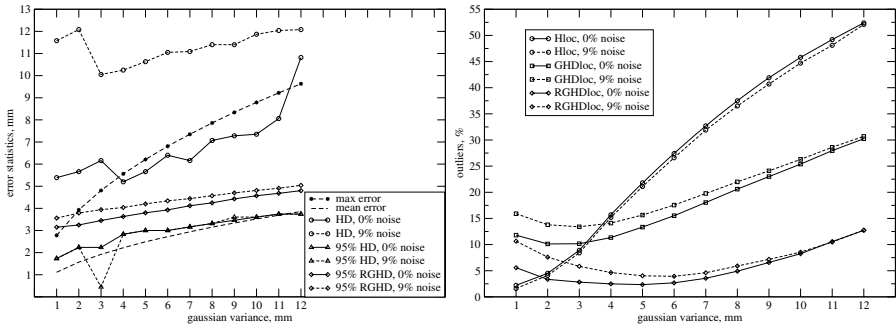
Fig. 3. Left: Error statistics for synthetically deformed BrainWeb images with and without noise, and the derived values of the Hausdorff distance based estimations. Right: The change in the proportion of the outlier measurements depending on the Gaussian variance.
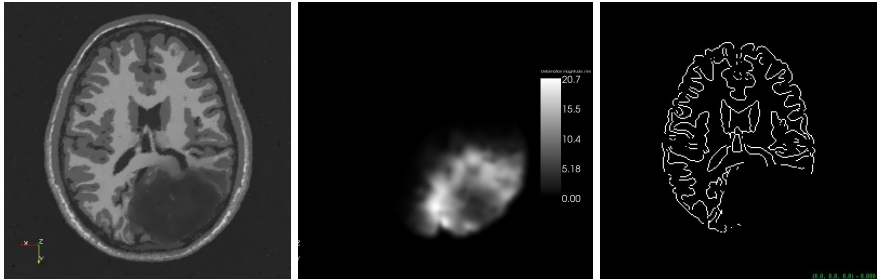
Fig. 4. Left: Synthetic tumor, case 2. Center: Deformation field produced by the tumor growth simulation (tumor mass effect and infiltration), colored by magnitude. Right: Edges recovered from the simulated tumor image. The same slice is shown in all images.

is consistently increasing as the alignment error increases, and it is always above the mean error value (see Figure 3, left). Thus, for large deformations (deformations as large as 10-15 mm have been reported during open scull craniotomy) RGHD is a more appropriate measure.

Synthetic tumor growth. We used simulated brain tumor growth images to assess error estimation performance for more realistic deformation modes, and for the images of different contrast. The images were created from the BrainWeb anatomical data as described in [18]. We used two versions of the simulated data: (1) with the intensity distribution close to that of the healthy subject image, and (2) with the intensity distribution derived from the real tumor data. Edge detection was done on the images with the regions corresponding to the tumor excluded. The misalignment was estimated between the healthy subject data and the image with the simulated tumor for the same subject at each feature point of the edge images. The recovered distances were compared with the true deformation magnitude from the tumor growth simulation (deformation field

Table 1. Outliers percentage for synthetic tumor growth data (only points corresponding to non-zero ground truth deformation are considered)

id	same contrast		diff. contrast		diff. contrast, enhanced	
	H_{loc}	RGH_{loc}	H_{loc}	RGH_{loc}	H_{loc}	RGH_{loc}
1	7.9%	4.6%	32.7%	42.9%	12.1%	9.5%
2	21.7%	16.4%	30.3%	34.1%	20.2%	15.4%
3	4.1%	3.6%	34.5%	44.7%	10.4%	7.1%

Table 2. Accuracy of error assessment for the tumor resection data in mm; empty entries correspond to image locations without edge features to assess error

landmark id	case 1			case 2			case 3		
	expert	H_{loc}	RGH_{loc}	expert	H_{loc}	RGH_{loc}	expert	H_{loc}	RGH_{loc}
1	2.59	–	–	2.51	–	–	4.43	1	3.07
2	0.48	–	–	1.52	–	–	4.53	1.41	2.25
3	0.48	1	0.92	2.99	1.41	1.53	3.96	–	–
4	0.48	1	1.2	1.36	1.73	1.16	2.15	–	–
5	2.59	1	0.82	0.98	–	–	2.88	2	2
6	1.07	–	–	2.4	–	–	3.66	–	–
7	2.45	–	–	2.04	–	–	3.49	2.24	2.22
8	1.44	1	0.63	1.92	1	1.36	4.43	1.41	2.94
9	3.36	2.24	3.45	3.04	–	–	3.96	–	–
10	1.44	1	1.11	1.36	1	1.43	1.98	1.41	1.05
avg difference w.r.t. expert		0.77	0.69		0.81	0.57		2.04	1.37

being the sum of the tumor mass effect and infiltration induced deformations). The outlier statistics is summarized in Table 1. Case 2 was the most complex, with the two infiltrating tumors of large volume located one next to another. Edge detection is particularly problematic in the edema region, which in this particular case extends to the majority of the deforming tissue region. This explains large number of outliers for set 2. Figure 4 helps to appreciate the complexity of error recovery for tumor set 2: there are very few edges detected in the area of the deformation, and the tumor area is almost indistinguishable from the large edema region. Nevertheless, robust HD estimation consistently has less outliers than the HD.

Neurosurgery registration data. We used three data sets from the public SPL repository of the tumor resection data[2]. Expert radiologist placed 10 corresponding anatomical landmarks in the pre- and intra-operative brain MRI T1 images. The error recovered using the HD-based techniques was compared with the expert-estimated error. The results are summarized in Table 2. On average, the RGHD measure shows better accuracy compared to the HD.

---

[2] http://www.spl.harvard.edu/pages/Special:PubDB_View?dspaceid=541
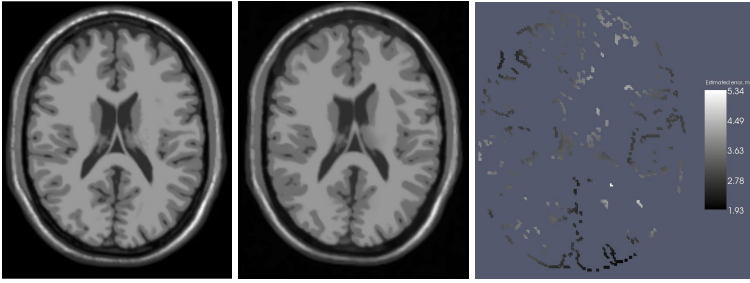
Fig. 5. Local estimation of misalignment using RGHD, all images show the same slice. Left: Undeformed image, BrainWeb. Center: Deformed image, Gaussian kernel variance 5 mm. Right: LDMap of the deformed and undeformed images, voxel values initialized to $RGHD_{loc}$.

4   Conclusions

We have presented an HD-based approach to estimation of image alignment error. Based on the evaluation results, the RGHD measure we propose can be more robust in terms of outliers in local distance estimation, and thus can potentially improve the accuracy of the image alignment assessment. While our primary application is the assessment of the non-rigid registration results, validation of the proposed method itself on real neurosurgery data is complicated by the absence of ground truth. Nevertheless, it can be used to improve the confidence in registration results.

The synthetic tumor growth data used in our evaluation may be more challenging than estimation of the pre-, intra-operative and registered image alignment. In the latter case, the images have similar content: tumor and edema are present on both images, and the edges detected from those images are more similar. We show that RGHD improves error estimation accuracy locally for anatomical landmarks, thus we expect that globally RGHD is also more accurate on neurosurgery data than the HD measure.

The evaluated techniques, and specifically RGHD – the most robust of the evaluated methods – can serve multiple purposes in registration assessment. First, they can be used as a global similarity metric between the two images, as well as for local alignment assessment. This mode of operation is particularly useful for automatic assessment of the non-rigid registration results during large-scale unsupervised parametric studies. Second, localized assessment of registration error can also be applied in conjunction with the visual assessment to provide quantitative error measurements. An example is shown in Figure 5. We emphasize, that the proposed method cannot substitute validation studies. Instead, it can be used in conjunction with other accuracy assessment methods for the patient studies, where accuracy is critical, processing time is highly limited and there are no means to compare the registration result with the ground truth. An promising area of our future work is the evaluation of the proposed measures in conjunction with the consistency tests of the deformation fields obtained during the NRR, and sensitivity of the measures to parameter selection of a specific NRR method.

References

1. Hill, D.L.G., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. Physics in Medicine and Biology 46, R1–R45 (2001)
2. Archip, N., Clatz, O., Whalen, S., Kacher, D., Fedorov, A., Kot, A., Chrisochoides, N., Jolesz, F., Golby, A., Black, P., Warfield, S.: Non-rigid alignment of preoperative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery. Neuroimage 35, 609–624 (2007)
3. Ferrant, M.: Physics-based Deformable Modeling of Volumes and Surfaces for Medical Image Registration, Segmentation and Visualization. PhD thesis, Universite Catholique de Louvain (2001)
4. Clatz, O., Delingette, H., Talos, I.F., Golby, A., Kikinis, R., Jolesz, F., Ayache, N., Warfield, S.: Robust non-rigid registration to capture brain shift from intra-operative MRI. IEEE Trans. Med. Imag. 24, 1417–1427 (2005)
5. Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H.: Introduction to the non-rigid image registration evaluation project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 128–135. Springer, Heidelberg (2006)
6. Archip, N., Tatli, S., Morrison, P., Jolesz, F., Warfield, S.K., Silverman, S.: Non-rigid registration of pre-procedural MR images with intra-procedural unenhanced CT images for improved targeting of tumors during liver radiofrequency ablations. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 969–977. Springer, Heidelberg (2007)
7. Huttenlocher, D., Klanderman, D., Rucklidge, W.: Comparing images using the Hausdorff distance. IEEE Trans. Pat. Anal. and Mach. Intel. 15, 850–863 (1993)
8. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 8, 679–698 (1986)
9. Lloyd, B., Szekely, G., Kikinis, R., Warfield, S.K.: Comparison of salient point detection methods for 3d medical images. In: 2005 MICCAI Open Source Workshop (2005)
10. Zhao, C., Shi, W., Deng, Y.: A new Hausdorff distance for image matching. Pattern Recognition Letters 26, 581–586 (2005)
11. Baudrier, E., Nicolier, F., Millon, G., Ruan, S.: Binary-image comparison with local-dissimilarity classification. Pattern Recognition 41, 1461–1478 (2008)
12. Peng, X., Chen, W., Ma, Q.: Feature-based nonrigid image registration using a Hausdorff distance matching measure. Optical Engineering 46, 057201 (2007)
13. Morain-Nicolier, F., Lebonvallet, S., Baudrier, E., Ruan, S.: Hausdorff distance based 3d quantification of brain tumor evolution from MRI images. In: Proc. 29th Annual Intl Conf of the IEEE EMBS, pp. 5597–5600 (2007)
14. Billet, E., Fedorov, A., Chrisochoides, N.: The use of robust local Hausdorff distances in accuracy assessment for image alignment of brain MRI. Insight Journal (January-June 2008), `http://hdl.handle.net/1926/1354`
15. Rousseeuw, P., Leroy, A.: Robust Regression and Outlier Detection. John Wiley & Sons, Chichester (1987)
16. Ibanez, L., Schroeder, W.: The ITK Software Guide 2.4. Kitware, Inc. (2005)
17. Rogelj, P., Kovačič, S., Gee, J.C.: Point similarity measures for non-rigid registration of multi-modal data. Comp. Vis. and Image Underst. 92, 112–140 (2003)
18. Prastawa, M., Bullitt, E., Gerig, G.: Synthetic ground truth for validation of brain tumor MRI segmentation. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 26–33. Springer, Heidelberg (2005)