

CHINESE MEDICAL SCIENCES JOURNAL

REVIEW

Accuracy Validation for Medical Image Registration Algorithms: a Review[△]

Zhe Liu¹, Xiang Deng², and Guang-zhi Wang^{1*}

¹Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China

²Center for Medical Imaging Validation, Corporate Technology,
Siemens Ltd. China, Beijing 100102, China

Key words: non-rigid registration; accuracy validation; digital phantom

Abstract Accuracy validation is essential to clinical application of medical image registration techniques. Registration validation remains a challenging problem in practice mainly due to lack of 'ground truth'. In this paper, an overview of current validation methods for medical image registration is presented with detailed discussion of their benefits and drawbacks. Special focus is on non-rigid registration validation. Promising solution is also discussed.

Chin Med Sci J 2012; 27(3):176-181

IN clinical applications, images acquired from multiple modalities, at different time or at different viewpoints need to be geometrically aligned for effective observation. Such a procedure of mapping points from one image to corresponding points in another image is called image registration. So far, many image registration algorithms have been proposed to address clinical needs. In a typical image registration process, an initial transformation model is constructed to align the moving image to the fixed image at first. Then the appropriate features are extracted from both images and the correspondences between these features are measured using 'similarity metric' to assess the quality of the align-

ment. At last the optimal transformation is estimated by minimizing the similarity metric using an iterative optimization method and embedded in a hierarchical scheme.¹

The registration techniques are generally categorized into two groups, rigid registration and non-rigid registration, according to the transformation model.² For rigid registration, only rotation and translation are considered and usually used in registering bone or brain images. For non-rigid registration, the correspondence between structures in two images can not be achieved without some localized warping. It is described by dense displacement fields defined at each voxel. Because most intrinsic deformation caused by patients' movement do not conform to a rigid approximation, the development of non-rigid registration techniques become main interest in medical image registration area nowadays.³⁻⁵

Validation is a very important post-step in registration algorithm development to make sure the clinical requirements can be met in the intended applications. The performance of the registration algorithm is typically evalu-

Received for publication April 25, 2012.

*Corresponding author Tel: 86-10-62783631, E-mail: wgz-dea@tsinghua.edu.cn

△Supported by National Basic Research Program of China Grant (2011CB707701) and National Natural Science Foundation of China (81127003).

ated using accuracy, robustness, stability, reliability, complexity, resources requirements and usability.⁶ Accuracy is probably the most important performance measure used to ensure the efficacy of registration algorithms in clinical use.^{7,8} However, it is not an easy task mainly due to the lack of absolute 'ground truth' in many clinical applications, especially those with non-rigid deformations.⁹⁻¹¹

CURRENT VALIDATION METHODS

Accuracy, however, can only be measured if a 'ground truth' transformation is available, which typically does not exist in clinical applications. In addition, there exists a paradox: if such 'ground truth' transformation is available, it should be used for the registration algorithm in the first place.¹² Solutions to the registration accuracy validation problem in previously published papers can be classified using the criteria shown in Table 1.

Visual inspection

Obviously, the most straightforward validation method is visual inspection by experts, although it seems informal and unreliable. In general, if the images look misregistered, they probably are misregistered indeed.¹² What's more, comparing the subtractions between fixed image to moving image before and after registration respectively can intuitively reflect the registration result.¹³ After all, qualitative visual inspection is rough and can only be used as a routine preliminary validation approach.

Cross-validation

Once a registration method is validated with some degree of accuracy, it can be used for cross-validating other methods.¹⁴ But it is not an optimal validation strategy in practice, because it is hard to reproduce the conditions in the original validation to ensure its validity. In addition, it is hard to say there is an authoritative registration algorithm for most clinical requirements. These problems may cause accumulated errors in cross-validation studies and give highly misleading results.

Landmark based validation

Since 'ground truth' transformation does not exist in

practical instances. Many investigators focus on a limited set of anatomical features for which homologies can be established between the image pair instead. For these features, comparisons can be made based on the obtained mappings or displacements between the images.¹⁵

A general validation technique is based on manual or automatic identification of anatomical landmarks in clinical data. In these researches, the locations of landmark points are labeled and the distance of corresponding points are calculated as validation metrics.^{16, 17} For example, the simplest metric is the average distance between corresponding point pairs:

$$M_{dist} = \frac{1}{m} \sum_{i=1}^m \|p_j - \hat{p}_j\|$$

where P_j and $\hat{p}_j$ is a corresponding point pair and $\|\cdot\|$ denotes the standard Euclidean norm.

Note that this validation method depends on the accuracy of the identification of homologous landmarks in two different images. So it is usually employed in registration of the vessels, muscle, bones and some other distinct structures.^{18, 19}

Segmentation based validation

The landmarks based validation method is based on the idea of representing the unknowable actual deformation by a group of points' displacement. But in some cases, e.g. liver image registration, reproducible identification of landmark points is difficult.

In this case, overlap of volume of interest to be registered can be an effective evaluation method.²⁰ The moving and fixed images were both manually segmented and labeled using interactive segmentation technique. Then several transformation models were applied to the corresponding moving labels. The resulting deformed labels were compared against the fixed labels by calculating the volume overlap, e.g.:

$$\text{Union overlap} = \frac{\sum_r |M_r \cap F_r|}{\sum_r (|M_r \cup F_r|)}$$

where M_r refers to a segmented region from registered moving image and F_r refers to the corresponding segmented region in the fixed image.

Table 1. Classification of the validation methods

Methodology	Ground truth	Data	Metric
Qualitative	Not used	Clinical data	Not used
Quantitative	Synthetic transformation	Physical phantom data	Target registration error
		Digital synthetic data	Target registration error
		Anatomical features	Average distance
	Not used	Clinical data	Volume overlap
			Consistency
			Intensity metric

Internal consistency based validation

An alternative approach is to examine the quality of registration without considering validation 'ground truth'. A very effective strategy for placing a boundary on registration accuracy is the application of internal consistencies.²¹ For an ideal registration, the forward transformation should be the inverse of the reverse transformation. This indicates that the registration is consistent. The cumulative inverse consistency error (CICE) is defined as the squared difference between the composition of the forward and reverse transformations and the identity mapping.

$$CICE_j(x) = \frac{1}{M} \sum_{i=1}^M \|h_{ji}(h_{ij}(x)) - x\|^2$$

where M is the number of images in the evaluation population, h_{ij} is the transformation from each evaluated image i to template image j .

Similarity metric based validation

Another intrinsic feature based validation method is to evaluate the registration results according to similarity metrics. Several intensity similarity metrics can be calculated from a registered image pairs for validation, including the root mean square of intensity differences (RMS_{int}), median-absolute deviation of intensity differences (MAD_{int}), and maximum intensity differences (MID_{int}).²² Note that no individual similarity metric is sufficient for an unbiased validation while comparing the algorithms using different similarity metrics because the similarity metric itself is implemented in the procedure of registration. However, evaluation using different similarity metrics will provide helpful information for validation.

Fiducial marker based validation

Applying fiducial markers on the physical phantoms or subjects then calculating the pair wise points' errors is the most employed method for validation of rigid registration. Rigid transformations of the fiducial markers attached to the anatomical structures during imaging process can be regarded as the 'ground truth'. A famous fiducial marker based validation project called 'retrospective inter-modality brain image' was proposed by West *et al.*²³ Users can download the marker-removed image *via* internet to apply their own registration algorithms and report the resulting transformations by the specification of the motion of eight corner points in the volume from which the transformation to be validated can be determined.

A more general validation application based on fiducial markers is calculating the target registration error (TRE). The registration accuracy at a 'target' point of interest which is not part of the set of registration points is

especially important in validation.²⁴ TRE is the distance from the true target position to the target position computed *via* the registration algorithm. The statistical distribution of TRE depends on the statistical distribution of the measurement errors, typically referred to the fiducial localization error (FLE):^{25, 26}

$$TRE(P)^2 = \frac{FLE^2}{N} \left(1 + \frac{1}{3} \left(\frac{d_x^2}{f_x^2} + \frac{d_y^2}{f_y^2} + \frac{d_z^2}{f_z^2} \right) \right)$$

where d_x is the distance of target point P from x axis, f_x^2 is the mean of the squared distances of the fiducial markers from that axis. So also does in y and z axes. And FLE can be obtained from fiducial registration error (FRE), which is the root-mean square of the distance between the corresponding marker points in two images.

$$FLE^2 = FRE^2 \times \frac{N}{N-2}$$

But in 2009, Fitzpatrick claimed that there is no correlation between FRE and TRE and using the value of FRE should be stopped. FRE is a reasonable indicator of whether or not a registration system is functioning properly but can not provide any information about the level of accuracy.²⁷

Digital simulation based validation

Another 'ground truth' transformation simulation strategy is to establish a digital model and deform it using appropriate spatial transformations with noise and other factors that can affect the registration accuracy.²⁸ Hipwell *et al.*²⁹ proposed a new validation method for X-ray mammogram registration algorithms, which uses 3D displacements obtained from computational bio-mechanical models of the breast.

However, digital simulations have definite limitations that can make them overestimate or underestimate registration accuracy. The validity of the validation is dependent on the simulation's realistic which may omit real factors that limit registration accuracy, or overestimate the influence of a limiting factor is actually present.

ACCURACY VALIDATION FOR NON-RIGID REGISTRATION

Although non-rigid deformation occurs much more frequently than rigid body deformation in reality, the development of non-rigid registration algorithms is impeded partly due to the difficulties in validation. Unlike the global rigid transformation, the non-rigid transformation is localized. That means some effective validation methods employed in rigid registration are not appropriate for non-rigid registration.

The conventional methods in non-rigid validation

Visual inspection, as a qualitative methodology, will be extremely misleading when evaluating the rigid registration algorithms, let alone the non-rigid ones with more degrees of freedom.

The quantitative validation techniques are mainly based on three different concepts according to the 'ground truth' situation.

1. Identify the location of corresponding anatomical features then compute their difference, such as landmark and segmentation based validation.

2. Measure the similarity metrics or consistency without considering the 'ground truth', such as consistency and intensity similarity based validation.

3. Simulate the 'ground truth' transformation based on synthetic data, such as fiducial marker and digital simulation based validation.

By using a large number of correctly identified anatomical features, such validation converges towards the ideal evaluation. But the major drawbacks of both segmentation and landmark based validation methods are inter-observer variability while manually identifying landmarks or labeling volume of interest. In addition, it is time consuming to satisfy the demand of sufficient landmarks or segments. Evaluation based only on a few images is not reliable for non-rigid registration.

Some researchers pay more attention to estimate uncertainty without 'ground truth'. These validations rely on real data and some metrics derived from the data. Besides the consistency and intensity metrics, some characterize the uncertainty of displacement field by calculating a covariance matrix which can be estimated according to bootstrap technique.³⁰ Such 'non ground truth' methods are clearly far from optimal, the validation result are not directly relevant to the accuracy of the alignment of images.

Fiducial marker stamped on physical phantom or patient body is the most effective and accurate validation method for rigid registration, however, it can not provide sufficiently dense displacement fields for non-rigid registration validation. What's more, physical phantom can not be deformed in a controlled manner to facilitate non-rigid transformation. Real patient body can provide realistic non-rigid deformation, but it is not feasible for all structures. For instance, neither external nor intrinsic markers can be stamped on the vessel trees.

Digital phantom for non-rigid validation

Digital simulation may give a clue that the spatial transformation model is likely to reflect the similarities of

modeling assumptions between the simulation and the registration method. One of the main limitations is that the simulated deformations are not necessarily representative of the range of deformations present in real patient data. To develop realistic and controllable phantom is a promising strategy for validation of non-rigid medical image registration.

Recently, more and more realistic and flexible phantoms are established which called hybrid phantom.³¹ The hybrid phantoms combine the realism of the patient-based voxelized phantom with the flexibility of the equation-based mathematical phantom. The most widely used hybrid phantom is extended cardiac-torso (XCAT) phantom which is the third generation of 4D cardiac-torso phantom series.^{32, 33} The original mathematical cardiac-torso (MCAT) phantom is based on abstract geometric primitives and then non-uniform rational basis spline (NURBS) surface is employed to represent anatomical dataset with higher level of realism (NCAT). After using higher resolution dataset and integrating NURBS surface with subdivision surface, the XCAT phantom can provide highly detailed whole-body anatomies at different respiratory time point. The simulated deformation in phantom can be used as 'ground truth' to validate registration algorithms.

Researchers have put much effort to improve the realism of digital phantom. Vandemeulebroucke *et al*³⁴ mentioned that the NCAT phantom remains a simplified, artificial version of the human thorax that can only approximate anatomical reality. As an improvement, they presented a 4D pixel-based and point-validated breathing thorax model (POPI), consisting of a 4D CT image data set along with associated vector fields and landmarks. It has been used as a 'ground truth' for validating non-rigid registration, for simulating motion compensated image reconstruction (CT, PET), and for computing 4D dose distribution in radiotherapy.

CONCLUSIONS

In this paper, we summarized current state-of-art accuracy validation methods for medical image registration techniques with an emphasis on non-rigid registration.

At present, there are still several major problems with validation of registration algorithms. First, absence of 'ground truth' transformation is still the major bottleneck in registration validation. Second, none of the validated metrics can determine whether the accuracy of an algorithm is good enough for clinical applications. Clinically relevant validation metrics need to be developed. Third,

validation methods are limited by their flexibility, and cannot be easily extended to validation of registration for other types of medical images.

In our opinion, construction of digital phantoms which can simulate the realistic deformation between two images is a more promising solution. With the development of free form modeling techniques, more and more hybrid digital phantoms have been proposed with synthetic deformation functions, such as the torso organs' respiratory motion³⁵ and brain tumor growth.³⁶ Digital phantoms not only help to quantify the spatial changes of patient's body, but are able to validate the registration algorithms for different types of medical images if integrated with imaging simulators.

In recent years, several voxel phantoms have been developed in China. The Chinese adult male voxel phantom is created using the color photographs of the Chinese Visible Human (CVH) datasets.³⁷ The voxel-based visible Chinese human (VCH) adult male phantom can provide high-quality details of anatomical characteristics of the Chinese population.³⁸ However, these voxel phantoms are developed for the validation of radiological dosimetry. The targeted hybrid phantom should be further improved, especially for the validation of medical image registration and other medical image processing algorithms.

REFERENCES

1. Klein S, Staring M, Murphy K, et al. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* 2010; 29:196-205.
2. Balasubramanian B, Porkumaran K. Registration of PET and MR images of human brain using normalized cross correlation algorithm and spatial transformation techniques. *J Theor Appl Info Technol* 2005; 16:1-8.
3. Balci SK, Golland P, Shenton M, et al. Free-form B-spline deformation model for groupwise registration. *Med Image Comput Comput Assist Interv* 2007; 10:23-30.
4. Metz CT, Klein S, Schaap M, et al. Nonrigid registration of dynamic medical imaging data using nD + t B-splines and a groupwise optimization approach. *Med Image Anal* 2011; 15:238-49.
5. Wang H, Dong L, O'Daniel J, et al. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys Med Biol* 2005; 50:2887-905.
6. Crum WR, Hartkens T, Hill DL, et al. Non-rigid image registration: theory and practice. *Br J Radiol* 2004; 77: S140-53.
7. Maintz JB, Viergever MA. A survey of medical image registration. *Med Image Anal* 1998; 2:1-36.
8. Hutton BF, Braun M. Software for image registration: algorithms, accuracy, efficacy. *Semin Nucl Med* 2003; 33: 180-92.
9. Wyawahare MV, Patil PM, Abhyankar HK, et al. Image registration techniques: an overview. *International Journal of Signal Process, Image Processing and Pattern Recognition* 2009; 2:11-27.
10. Rueckert D, Sonoda LI, Hayes C, et al. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* 1999; 18:712-21.
11. Scheys L, Loeckx D, Spaepen A, et al. Atlas-based non-rigid image registration to automatically define line-of-action muscle models: a validation study. *J Biomech* 2009; 42:565-72.
12. Schnabel JA, Tanner C, Castellano-Smith AD, et al. Validation of nonrigid image registration using finite-element methods: application to breast MR images. *IEEE Trans Med Imaging* 2003; 22:238-47.
13. Periaswamy S, Farid H. Elastic registration in the presence of intensity variations. *IEEE Trans Med Imaging* 2003; 22:865-74.
14. Heinrich MP, Schnabel JA, Gleeson FV, et al. Non-rigid multimodal medical image registration using optical flow and gradient orientation. *Proc of Medical Image Understanding and Analysis (MIUA)*. Coventry: Department of Computer Science, University of Warwick, 2010.
15. Rogelj P, Kovacic S, Gee JC. Validation of a non-rigid registration algorithm for multi-modal data. *Proc SPIE* 2002; 4684:299-307.
16. Choe AS, Gao Y, Li X, et al. Accuracy of image registration between MRI and light microscopy in the *ex vivo* brain. *Magn Reson Imaging* 2011; 29:683-92.
17. Lange T, Lamecker H, Hunerbein M, et al. Validation metrics for non-rigid registration of medical images containing vessel trees. *Proc Workshop Bildverarbeitung für die Medizin (BVM)*. Berlin: Springer-Verlag, 2008.
18. Markelj P, Likar B, Pernus F. Standardized evaluation methodology for 3D/2D registration based on the Visible Human data set. *Med Phys* 2010; 37:4643-7.
19. Ireland RH, Dyker KE, Barber DC, et al. Nonrigid image registration for head and neck cancer radiotherapy treatment planning with PET/CT. *Int J Radiat Oncol Biol Phys* 2007; 68:952-7.
20. Klein A, Ghosh SS, Avants B, et al. Evaluation of volume-based and surface-based brain image registration methods. *Neuroimage* 2010; 51:214-20.
21. Christensen GE, Geng X, Kuhl JG, et al. Introduction to the non-rigid image registration evaluation project (NIREP).

- Proc SPIE 2006; 4057:128-35.
22. Yin LS, Tang L, Hamarneh G, et al. Complexity and accuracy of image registration methods in SPECT-guided radiation therapy. *Phys Med Biol* 2010; 55:237-46.
 23. West J, Fitzpatrick JM, Wang MY, et al. Comparison and evaluation of retrospective inter-modality brain image registration techniques. *J Comput Assist Tomogr* 1997; 21:554-66.
 24. Segner A. Rigid-body point-based registration: The distribution of the target registration error when the fiducial registration errors are given. *Med Image Anal* 2011; 15:397-413.
 25. Fitzpatrick JM, West JB, Maurer CR. Predicting error in rigid-body point-based registration. *IEEE Trans Med Imaging* 1998; 17:694-702.
 26. Tomazevic D, Likar B, Pernus F. "Gold standard" data for evaluation and comparison of 3D/2D registration methods. *Comput Aided Surg* 2004; 9:137-44.
 27. Fitzpatrick JM. Fiducial registration error and target registration error are uncorrelated. *Proc SPIE* 2009; 7261: 7261021-12.
 28. Rexilius J, Hahn HK, Schluter M, et al. Evaluation of accuracy in MS lesion volumetry using realistic lesion phantoms. *Acad Radiol* 2005; 12:17-24.
 29. Hipwell JH, Tanner C, Crum WR, et al. A new validation method for X-ray mammogram registration algorithms using a projection model of breast X-ray compression. *IEEE Trans Med Imaging* 2007; 26:1190-200.
 30. Kybic J. Bootstrap resampling for image registration uncertainty estimation without ground truth. *IEEE Trans Image Process* 2010; 19:64-73.
 31. Johnson PB, Whalen SR, Wayson M, et al. Hybrid patient-dependent phantoms covering statistical distributions of body morphometry in the U.S. adult and pediatric population. *Proc IEEE* 2009; 97:2060-75.
 32. Segars WP, Tsui BM. MCAT to XCAT: the evolution of 4D computerized phantoms for imaging research. *Proc IEEE* 2009; 97:1954-68.
 33. Segars WP, Sturgeon G, Mendonca S, et al. 4D XCAT phantom for multimodality imaging research. *Med Phys* 2010; 37:4902-15.
 34. Vandemeulebroucke J, Sarrut D, Clarysse P, et al. The POPI-model, a point-validated pixel-based breathing thorax model. *Proc. of the 15th International Conference on the Use of Computer in Radiation Therapy (ICCR)*. Oakville: Novel Digital Publishing, 2007.
 35. Dey J, Pan T, Choi DJ, et al. Estimation of cardiac respiratory-motion by semi-automatic segmentation and registration of non-contrast-enhanced 4D-CT cardiac datasets. *IEEE Trans Nucl Sci* 2009; 56:3662-71.
 36. Bauer S, Seiler C, Bardyn T, et al. Atlas-based segmentation of brain tumor images using a Markov random field-based tumor growth model and non-rigid registration. *Conf Proc IEEE Eng Med Biol Soc* 2010; 2010:4080-3.
 37. Zhang B, Ma J, Liu L, et al. CNMAN: a Chinese adult male voxel phantom constructed from color photographs of a visible anatomical data set. *Radiat Prot Dosimetry* 2007; 124:130-6.
 38. Zhang G, Luo Q, Zeng S, et al. The development and application of the visible Chinese human model for Monte Carlo dose calculations. *Health Phys* 2008; 94:118-25.