A Survey on Evaluation Methods for Medical Image Registration

Ximiao Cao , Qiuqi Ruan
Institute of Information Science, Beijing Jiaotong University, Beijing, China

Abstract - **How to estimate the results of medical image registration is still a problem, because no "golden estimation criterion" has been proposed. This paper not only presents an overview of existing estimation criterions for medical image registration, including the advantages and shortcomings of each estimation method, but also proposes some improved methods for estimation. At last, it gives out a summary of some problems existing in this field, which may be the hot in the future.**

Keywords - **Medical image registration, registration evaluation, method characterization, performance assessment.**

I. Introduction

Registration of medical images is an essential process in medical image processing applications, especially for image fusion. The purpose of registration is to provide exact evidence for treatment. So far, many image registration algorithms have been proposed based on different image features. Generally when the algorithms are estimated, more attention is paid on precision, stability, reliability, complexity and usability, especially on the first two parameters [1]. In Part Ⅱ, more attention is paid on the results of different algorithms. Part Ⅲ pays more attention on estimating the character of the algorithms. Part Ⅳ proposed the means used in papers.

II. Estimation by Registration Result

Early, E. Walter proposed the estimation criteria of SSC and DSC [2] (Stochastic and Deterministic Sign Change Criteria). The nonparametric approach was demonstrated to outperform the conventional image registration criteria for robust registration.

Some years later, Chiang and Sullivan [3] compare a new criterion for image registration called CBC (Coincident Bit Counting) with two criteria of SSC and DSC. The CBC method compares the number of coincident bits between the corresponding pixels in two different frames for a fixed amount of displacement.

Recently, many new approaches are proposed while some new methods on estimating the results of matching come out.

A. References Subjective Estimation by Experts

At present, the subjective method mainly refers to the estimation given out by the therapists [4]. Both original and matched medical images are presented separately to several therapists. It is testified that the discrimination could achieve to one pixel by eye-estimating [5]. The grades of results are as follows:

1) *Good:* Good enough to be used, the error is about one pixel distance.
2) *Accepted:* Accepted but not as well as (1). The error is bigger than one pixel.
3) *Not Accepted:* The results are not good enough to use. The error is more or less bigger than one pixel, but less than three pixels.
4) *Bad:* the error is bigger than 3 pixels.

The objective method is easy to operate, and the grading is according to the clinical practicality. However, the standard is objective and can not be quantified. The same image could be estimated as different grading. Moreover, such estimation lacks statistical testing, because only part of the registration results could be estimated, whose results represent the veracity of the algorithm. In a word, this method of estimation is direct and effective when the golden one has not been found.

B. Estimation by models

This is an indirect estimating method.

Timothy [6] simulated head by a cylinder full of water, with imaginable marks on proper position. Ge [7] made the nails in skull as identifiers in CT/MRI/PET, and estimated the veracity of algorithm by the distance of markers in the images after matching. Evans [8] simulated PET images by MRI images, and tested the registration results after processing images by rotation, moving, resizing, and adding noises. Visible Human CT is usually used as reference image because of its high definition, clear configuration and clear information of location.

The distance between the marking points in the matched image can be shown accurately, and such estimation can be quantified. However, the estimation is still far from the real case, especially when simulating a non-rigid object [9].

C. Estimation by mutual information

The criterion of mutual information is effective, especially when aiming at the methods based on mutual information. However the application is limited because it demands the grey degrees have the corresponding relationship in the base image and input image. (Base image: this is the image against which you compare the image to be registered. It is also often called the reference image. Input image: this refers to the image that you wish to register. It is often called the observed image)

The paper [10] proves the effectiveness of algorithms by comparing the mutual information before matching with that after matching.

In this paper, a parameter of "matched information rate" is proposed.

$$\rho = 2\times I(X;Y)/[H(X)+H(Y)] \quad \rho \in [0,1] \tag{1}$$

This parameter is better in demonstrating the degree of matching. According to the definition of relativity in the theory of information, when the base image and the input image are non-relative, the mutual information is zero, and ρ is zero. When the output image is the best result, the mutual information and the parameter ρ achieve the maximum. If the base image and the input image are the same, the better the images registering, the parameter ρ is greater. The parameter achieves to 1 when input image is the same with the base image, and they are totally matched.

D. Estimation by rate of good matching

In paper [11], a MR-T2 medical image with is selected, and is processed by rotation and moving to get another 30 images. The moving distance is generated randomly in [-10, 10] (mm), while the angle rotated is [-5°, 5°]. Matching the original image with the 30 images for 30 times, it will be regarded as a wrong matching when the angle error is bigger than 1°or the distance error is bigger than one pixel. The performance is evaluated by calculating $\eta_{success}$ as:

$$\eta_{success} = n_{matched} / n_{total} \tag{2}$$

Thus, $\eta_{success}$ is a very important parameter, for $\eta_{success}$ is larger while the introduced error may be less and the algorithm is better. The advantage of this method is its explicit and practical meaning. Unfortunately, this method can not be adapted when estimating non-rigid image registration and multimodality image registration, because it only pays attention on the spatial information.

III. Estimating the character of algorithms

The main problems in robustness of the algorithms are as follows:

1) Not enough effective information is contained in the test image, e.g. registration of head images.

2) Deformation in images, e.g. when matching the abdominal images of MR and PET, the main reason for wrong matching is for the deformation in PET.

3) To minimize the quantity of data in the algorithm based on MMI, the definition and information in image will be reduced, which will affect the performance.

4) The noise in images reduces the relativity between the test and reference images.

A. Analyzing curves of MMI

According to the reason above, in paper [12], the performance about robust is evaluated by calculating the MMI at different transformation. The curve shows the local maximum clearly. If the local maximum could be avoided, the robust character of the algorithm is well resolved.

B. Analyzing character of anti-noise

Add some noise into the test image, then register this image with the reference image, and compare this result with that without noise adding. If the results are almost the same, the algorithm has a good performance in anti-noise. In paper [13], pepper noise was bought in to test the anti-noise character. In paper [14], white Gauss noise with means of zero, variance of 0.01, 0.05, 0.1, and 0.2.

C. Analyze character of practicality

A good algorithm should be well adapted on different people. In order to test the practicality of the algorithm, it is necessary to have tests on images from different patients. Moreover, to compare two algorithms precisely, it is better to test the algorithm on the same image resource. That is, a standard database of medical image is of great importance and necessary when estimating. National Library of Medicine provides some standard medical image for worldwide researchers.

D. Estimating the segmentation methods

Segmentation is a basic step in some algorithms of image registration. The performance of this step will have great affluence on the results of the registration. Thus, how to estimate the performance of the step of segmentation should be paid high attention. In paper [9], four parameters are provided:

1) Contrast between regions, which could be used to estimate the quality of segmentation.

2) Uniformity in regions.

3) Constringent probability which figures out the stability of algorithms, and diffusion coefficient which figures out the consistency of the algorithms.

4) The mistake rate which figures out the number of pixels that are divided incorrectly.

IV. Parameters for estimation

A. Quantity of calculation

The final purpose of algorithm research is to find a swift and effective method for medical application. Thus, the cost of the algorithms including time cost and the memory cost is a significant index. Moreover, when given out the time cost, the hardware condition of the computer should also be given.

B. Error parameter

The error includes the angle error and the moving error, while the formation of error contains mean error, maximum error, variance, and error median. In paper [15], mean error, maximum error and variance of moving error of X and Y label and angle error are used. In paper [16], mean error, maximum error as well as median error is used to reflect the performance of registration.

Error parameter is calculated by:

$$E = \sum_i [A'(x_i, y_i) - B(x_i, y_i)]^2 = \sum_i e_i^2 \tag{3}$$

$A(x_i, y_i)$ is the test image, while $B\ (x_i, y_i)$ is the reference image; $A'(x_i, y_i)$ is the result of transforming $A(x_i, y_i)$. $A'(x_i, y_i)$ and $B'(x_i, y_i)$ are the corresponding mark pair.

C. Correlation index

Classical calculation on correlation includes self-correlation, unitary correlation, statistical correlation, and phase correlation. The formula is as follows:

$$Similar\ (X_1, X_2; Y_1, Y_2) = cor\ (X_1, X_2) \bullet \rho(X_1, X_2; Y_1, Y_2) / [1 + dist\ (X_1, X_2; Y_1, Y_2)] \quad (4)$$

Where

$$dist\ (X_1, X_2; Y_1, Y_2) = d\ (X_1, Y_1) + d\ (X_2, Y_2) / 2$$

$$d\ (X, Y) = \| X - Y \|$$

$$\rho(X_1, X_2; Y_1, Y_2) = \begin{cases} e^{\frac{-r}{\delta}}, & \text{if } (Y_1, Y_2) \text{is desired match pairs, and } r < \delta \\ 0, & \text{others} \end{cases}$$

$$\delta = 0.3$$

$r = [d\ (X_1, Y_1) + d\ (X_2, Y_2)] / dist\ (X_1, X_2; Y_1, Y_2)$, which is the relative distance.

The correlation index is based on the relative position (the value of r). Every desired point (Y_1, Y_2), which to be matched with (X_1 , X_2), contributes as an exponential attenuation function. While r is larger, $\exp(-r/\delta) \to 0$, the desired point is neglected; While $r \to 0$, $\exp(-r/\delta) \to 1$, the desired point (Y_1, Y_2) contributes much to the correlation index of (X_1, X_2).

D. The degree of matching

Every feature point in test image has several desired points to be matched with in reference image. Also every feature point in reference image has several desired points to be matched with in test image. Select one of the desired points with maximum correlation index. The degree of matching is calculated by the formula below.

$$Strength\ (X_1, X_2) = \sum_{Y_1 \in N(X_1)} \max_{(Y_1, Y_2)} [Similarity\ (X_1, X_2; Y_1, Y_2)] + \sum_{Y_2 \in N(X_2)} \max_{(Y_1, Y_2)} [Similarity\ (X_1, X_2; Y_1, Y_2)] \quad (5)$$

In paper [17], the degree of matching is used to estimating algorithms, which is testified to be a continuous and veracious method for estimation.

IV. Estimation in application

The methods and parameters mentioned above are relative, not unattached. In actual, several methods and parameters are used together for estimating the algorithms from all sides. On the other hand, not all the parameter should be calculated for one algorithm, because of their overlapped practical meaning. How to select the parameters to demonstrate the capacity from all sides is of great importance when estimating an algorithm.

In paper [18], the author adopts both subjective and objective method to estimating the algorithms. Firstly, hand in the fusion image with test and reference images (PET and CT) to the therapists, who will give out estimation objectively, in the meaning of clinical. Secondly, extract out the boundary in the two images, calculate the error between which gives a quantity to the location difference.

In paper [19], the time cost is given out when analyzing the effectiveness of the algorithm. At the same time, the paper also gives out the clinical estimation to prove the practicality of the algorithm.

In paper [20], the curves about the matching parameters were firstly given out to estimating and comparing the two algorithms. The sharper the curve is, the easier to find out the optimal matching. Secondly, an experiment testing the affection of grey levels was given by calculating the matching parameters under the condition of both full grey levels and reduced grey levels. The affection is quantified by the mean error between those under the two conditions above. Thirdly, the character of anti-noise of the algorithm is estimated by adding white Gauss noise with zero mean and variances of 0.01, 0.05, 0.1 and 0.2 into five groups of image. Then draw out a table of the errors of matching parameters under the four conditions to reflect the anti-noise character of the algorithm. This character is finally estimated by the maximum of error, after calculating the mean of five groups under different conditions. At last, the objective estimation method is also used.

Thus, this paper presented an all-round approach method to estimate the character of the algorithm as well as to analyze the result of output image.

In paper [14], fifty times of tests are made to testify the effectiveness of the algorithms by the parameter of the percentage of successful matching in all. Also, estimating parameters such as maximum MMI, average MMI, the shortest time cost and the average time cost, as well as the times of matching that could achieves sub-pixel degree, are also given out.

In paper [17], the algorithm is analyzed by the shape of parameter curves. Meanwhile, the character of anti-noise is also tested by add Gauss noise.

V. Conlcusion

At present, there are still several problems existing on estimating methods.

1) *Limited effectiveness*

The different estimating methods used now are effective to some extent, especially when comparing two algorithms of matching. However, there is still a long way to make good use of these methods and parameters to find a way to estimate an algorithm from all sides. Researchers have not found an excellent method or a standard process to analyze the matching algorithms. Parameters and methods have their limitedness and shortcomings. For example, the application of MMI method for estimation is limited because it demands the grey degrees have the corresponding relationship in the base image and input image.

2) *Quantify the parameter*

Several quantified parameters have been used to estimate algorithms. However, there has not been a standard and authorized value of parameter from clinical that could determine whether the algorithm is good enough for practicality. That is, the relationship between the quantified parameter and the clinical meaning is still confused.

3) *Resource of image*

Creating a standard resource of medical images is necessary and significant. When comparing algorithms based on medical images, it is more convincible if they are tested on the same resource of images. Also, we could remove the affluence on image from different equipment, the different definition and grey levels, the different process of imaging, and so on.

It is ineluctable to compare one algorithm with another, as more and more new algorithms of medical image matching are proposed. In order to find an ideal algorithm for clinical use, estimation should be paid more and more attention. There is still a long way to find out the "golden standard", which is also our next step.

References

[1] West J., "Comparison and evaluation of retrospective inter modality Image registration technique [J]," *Medical Imaging: Image Processing* 1996, 2710:332-347.

[2] E.Walter, L.Pronzato, A.Venot, "Theoretical properties of sign change criteria for robust off-line estimation" *Automatic*, vol.25, pp.949-952, 1989

[3] J.Y. Chiang, B.J.Sullivan, "Coincident bit counting—A new criterion for image registration," *IEEE Trans, Med Imaging*, vol.12, pp.30-38, 1993

[4] QIN Bin-jie, ZHUANG Tian-ge, "Multi-resolution registration of MR and CT images based on correlation ratio similarity measure," *Chinese Journal of Biomedical Engineering,* Vol.22 No.1 February 2003

[5] Studholme, C, Hill DLG, Hawkes DJ. "Automated 3-D registration of MR and CT images of the head [J]" *Med Image Analysis*, 1996,1 (2): 163-175.

[6] Timothy GT, Ronald JJ. "Accuracy of registration of PET, SPECT and MR images of a brain phantom [J]" *Medical Imaging: Image Processing,* 1996, 2710:332-347

[7] Ge, Y, Fitxpatrick JM. "Retrospective registration of PRT and MR brain images: an algorithm and its stereo tactic validation [J]". *Compute Assistant Tomograph*, 1994, 18:800-810.

[8] Evans AC, Collins DL, Nellin P, et al. "Correlative analysis of three imensional brain images computer – integrates surgery [J]" *Technology and, Clinical Applications*, 1996, 99-144.

[9] ZHANG Jian-ge "Overview on Medical Image Registration" *Academic Journal of Shanghai Second Medical University* 0258-5898(2003)02-0180-03

[10] ZONG Ren-he, ZHU Xiang-sheng, LIU Liang-cheng, ZENG Gang-yan "Mutual Information Based Medical Image Registration" *Computer Technology and Application Development2004*

[11] GAO Zhi-yong, GU Bin, LIN Jiarui, "Implementation of Mutual Information Based Medical Image Registration Methods" *Journal of Biomedical Engineering* 2003:20(3):476-478

[12] CHEN Ming, CHEN Wu-fan, FENG Qian-jin, YANG Feng, "Medical Image Registration Based on Mutual Information and Fuzzy Gradient Similarity" *ACTA ELECTRONICA SINICA* Vol.31, No.12, Dec.2003, 0372-2112(2003) 12-1835-04

[13] YANG Feng-bao, YANG Fang-lin, WEI Quan-fang, HAN Yan, "Study on Image Registration Algorithm Based on Image Trace", *Journal of Test and Measurement Technology,* Vol.20, No.1, 2006, 1671-7449(2006)01-0071-04

[14] FENG Lin, YAN Liang, HUANG De-gen, HE Min-feng, TENG Hong-fei, "A Study of PSO and Powell Hybrid Algorithm Applied in Medical Image Registration", *Beijing Biomedical Engineering,* Vol.24, No.1, Feb. 2005: 1002-3208 (2005) 01-0008-05

[15] PENG Jing-lin, ZHANG Jing, LI Shu-tao, WANG Lian-hong, "Medical Image Registration Based on Expanded PV Interpolation and Hybrid Optimization Algorithm", *ACTA ELECTRONIC SINICA* Vol.34, No.5 May, 0372-2112(2006) 05-0962-04

[16] ZHOU Yong-xin, LUO Shu-qian, "Medical Image Registration Based on Mutual Information of Feature Points", *Journal of Computer-Aided Design & Computer Graphics*, Vol.14, No.7, July, 2002 TN919.8

[17] ZHAO Yong-ming, ZHANG Su, ZHANG Zhao, CHEN Ya-zhu, "A Non-rigid Medical Image Registration Algorithm Based on Modally Controlled Free Form Deformation", *Space Medicine & Medical Engineering*, 240-245, Vol.18, No.4, Aug.2005

[18] YI Chuan-liang, DU Ming-hui, "CT/PET Image Registration Based on a Region Similarity Measure", *Chinese Journal of Medical Instrumentation*, Vol.29, No.1, Sep. 2005: 1671-7104(2005)01-0007-04

[19] HAI Miao, SHEN Yi, "Research on automatic registration method for 3D multi-modal images" *Chin Med Imaging Technol*, 2005, Vol.21, No.5

[20] LI Ke, LIU Chang-chun, LI Tong-lei, "Medical registration approach using improved maximization of mutual information", *Journal of Shandong University (Engineering Science),* Vol.36, No.2, Apr. 2006, 1672-3961(2006)02-0107-04