Internal Evaluation of Registration Results for Radiographic Images

Boris Peter Selby[1]*, Georgios Sakas[2], Stefan Walter[1], Wolf-Dieter Groch[3], Uwe Stilla[4]

[1] *Medcom GmbH, Rundeturmstr. 12, 64283 Darmstadt, Germany*
[2] *Fraunhofer IGD, Cognitive Computing and Medical Imaging, 64283 Darmstadt, Germany*
[3] *University of Applied Sciences, Fachbereich Informatik, 64295 Darmstadt, Germany*
[4] *Technische Universitaet Muenchen, Photogrammetry and Remote Sensing, Germany*

Abstract

This work focuses on internal gray level based evaluation of image registration results. The motivation is to provide an approach for self-diagnosis in the scope of a patient alignment system based on rigid registration of real and reconstructed X-ray images. As an automatic system should provide expressive indicators for the correctness of the outcome, we propose a method to estimate the probability for the resulting transformations to lie within a predefined window of acceptable values. Based purely on image gray values, the approach is independent from previous knowledge about the images. By registration of corresponding fragments of both images we generate redundancy and define the probability density of the resulting transformations. The proposed method is tested comparing digital reconstructed radiographs (DRRs) to X-ray images. By introducing geometric and radiometric deviations we show that a reliable self-diagnosis is possible.

1. Introduction

Algorithms for image registration have been a matter of intensive research over the past few years and are used in a large range of applications. In many cases, if images from different sources are to be registered, feature based algorithms can hardly be used to find the transformation between the images. Therefore gray level based approaches as Mutual Information (MI) are widely spread. Their advantage is that they do not depend on detection of corresponding image features and are applicable to a variety of image modalities. A major drawback is that the image similarity measure maximized by registration depends on image properties [1] and a reliable evaluation of the quality of the result is not possible by means of the similarity measure.

The motivation for our approach is an application resided in image guided radiotherapy (IGRT). Particle beam therapy allows accurate application of the radiation dose onto the diseased tissue with accuracy less than 1.0 mm and efforts accurate patient alignment [2]. Image guided approaches allow automatic alignment of patients by comparison of digital radiographs (DRs) and DRRs. The DRRs are reconstructed from CT data and rigid registration of DRRs with DRs gives the alignment error of the patient in image space, which can be back-projected into 3D space to realign the patient [3, 4]. However, an automatic system should provide the user with meaningful values indicating the reliability and accuracy of the obtained results. This can be referred to as self-diagnosis [5].

The crucial factor for the alignment process is the automatic image registration part. Thus, we focus on the registration result and propose a general approach for internal evaluation of multi-modal image registration, which could also be deployed in the scope of other registration applications.

2. Related Work

Many suggestions have been made for external evaluation of registrations. E.g. in [6] a method is proposed to evaluate different image registration methods based on previously known displacements. Unfortunately these methods are not applicable here, as they assume that the real misalignment of the imaged object is known in advance – thus it is an external evaluation.

There are also a couple of publications dedicated to the problem of internal evaluation of image matching or interpretation results. In [7] a self-diagnosis method is proposed for detection of roads in aerial images. Assuming, that if some content of corresponding images can be interpreted, it should certainly be possible to find a transformation to match these. The proposed method is based on an underlying model combining simple image features to objects of higher order. It is possible to define semantic and geometric properties that hold with all images. This approach is hardly applicable for medical images gained from different sources, for different body parts and varying viewing angles, especially in our case where X-ray images are used, not showing a solid surface of an object but the line integral of rays.

The work of [8] focuses on a neural network technique to analyze the registration quality for rigid registration using stochastic moments of the image histograms. However, neural networks come with significant drawbacks. They must be trained with respective input data. In case of the proposed method, only DRs of the head and neck area were used. Another problem of neural nets is, that it is hard to derive meaningful values from the results, e.g. if the net assigns some quality value to a registration, the major questions remain: How accurate was the registration, and with what probability?

In our approach we produce redundant image registration results and interpret these statistically to derive the probability with which the registration result lies in a predefined window of tolerated error.

3. Methods

There are two possibilities to perform an internal evaluation: a) by using additional information, not involved in the registration process or b) by using redundant information. In our case we are not provided with any additional information but the image data itself. Thus we use redundant registration results for the evaluation.

The approach can be described in three parts. First two registered images are subdivided into single fragments, which are registered independently. Then the relative histograms are built from the resulting transformations and outliers are suppressed. In the last step, the histograms are combined to a single n-dimensional probability density function (pdf) where n denotes the number of histograms or degrees of freedom, given from the registration. From the pdf the probability to fall in the range of a tolerated displacement is computed.
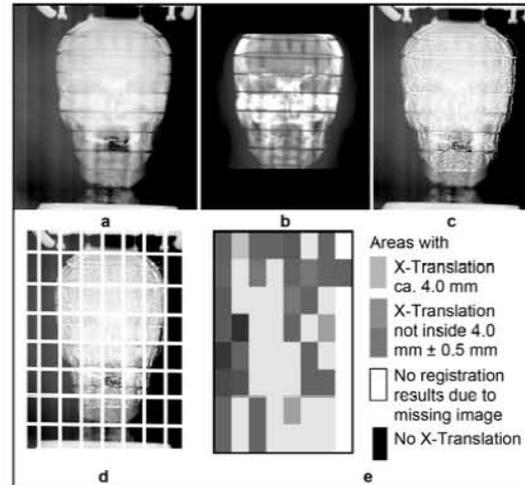

Figure 1. Images of a head phantom a) DR; b) DRR; c) DR overlaid with contours from DRR, displaced by 4.0 mm; d) 8x8 fragments for registration; e) Displacement along image X-axis for single fragments after registration

3.1. Registration of fragmented images

First the registered images are divided into 8x8 pairs of rectangular fragments. We use this number dependent on the DRR image resolution of 512x512 pixels, assuming that a total number of 64x64 pixels in one fragment should be enough to perform a registration (Fig. 1a-d).

Each fragment pair contains a region of the reference image (A = X-ray) and the floating image (B = DRR). If the images are registered properly no transformational offset can be found and the transformation T, mapping one fragment to another, is the identity matrix. The single regions are registered using MI as similarity measure [1], computed by equation 1, where $H(A)$ and $H(B)$ denote the entropies of the respective image-fragments and $H(A, TB)$ is the joint entropy:

$$MI(A,B,T) = H(A) + H(B) - H(A, T*B) \qquad (1)$$

We register the images by minimization of the negative MI. In our case the transformation T defines three degrees of freedom (x-shift tx, y-shift ty and image plane rotation rz). This is because the final evaluation result shall take these degrees of freedom into account. If e.g. an additional scaling between the images has to be evaluated, one could simply add this degree of freedom to the single registrations.

Minimization is done by a Downhill Simplex (DHS) approach as described in [9]. To avoid getting

stuck in local minima of the MI function the DHS optimizer is modified. When a minimum is found the algorithm is repeated, starting from the optimized transformation values with an enlarged simplex. This is done until no smaller negative MI value can be found. This approach can be considered a mixture of Downhill Simplex and Simulated Annealing [9].

Finally the three transformation values for each fragment are obtained. They may vary for each fragment due to the different appearance of the images (Fig. 1e). The single components of the fragment transformations are stored as n-vectors tx_n, ty_n and rz_n. Some fragments cannot be registered because they are defined in image regions that do not overlap. So n denotes the number of successful registrations with $n \leq$ *Number of Fragments*.

3.2. Relative Histograms of Transformations

For each degree of freedom from the registration, a relative histogram (h_{tx}, h_{ty} and h_{rz}) is built, giving the probability for the occurrence of a certain transformation. The histogram bin size is 0.5 pixels for the shifts. This is the expected rounding error when computing the translations, as nearest neighbor interpolation is used in the registration. For the rotation histogram we use 0.1° bins.
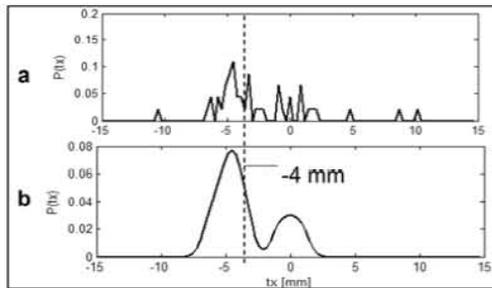

Figure 2. a) Relative histogram of the x-translations, -4 mm image misalignment; b) Outliers removed and curve smoothed

If the images differ in their appearance, the histograms can contain outliers with low probability (see Fig. 2a). The median of the probabilities is used to remove these. Large probabilities are preserved by using the fraction of $k = 0.25$. For transformations having a probability $P(t_i) < P_{median}$ the probability is set to $P(t_i) = 0$. The histogram is rescaled so that the sum of probabilities is still *1*. Obviously erroneous transformations resulting from improper registration of the image fragments can be removed a priory by this process.

To finally smooth the histograms the Parzen Window technique is used [10]. The Parzen window function is the pdf for normal distributed random values. For each histogram entry a smoothed value is estimated by equation 2:

$$P(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2\pi\sigma} \exp\left(-\frac{(t_i - t)^2}{2\sigma^2}\right) \quad (2)$$

Where n is the number of single probabilities and is set to the size of the histogram bin. The smoothed result can be seen in figure 2b.

3.3. Evaluation of the Histograms

The last step is to combine the histograms to obtain a single value that can be used as quality indicator for the registration. Given the pdfs of the histograms the expected value, the variance and the co-variances are computed for each histogram and therewith for every transformation type.

For the possible transformation types tx, ty and rz we determine the 3 x 3 co-variance matrix and define a multivariate pdf f_t in three dimensions by

$$f_t = \frac{1}{2\pi^{n/2}\sqrt{\det C}} \exp\left(-\frac{1}{2}(t-\mu)^T C^{-1}(t-\mu)\right) \quad (3)$$

Where n is the number of parameters to evaluate (here 3), t is a 3-vector containing three specific transformations and is the 3-vector of the expected values. One can now define a range of accepted parameters, e.g. $TX = \pm0.5\ mm$, $TY = \pm0.5\ mm$ and $RZ = \pm0.5°$. From the pdf in equation 3 the probability for which the registration result lies inside the given range is determined by integration over the parameters t:

$$P(TX, TY, RZ) = \int_{-TX}^{+TX} \int_{-TY}^{+TY} \int_{-RZ}^{+RZ} f(tx, ty, rz)d_{rz}d_{ty}d_{tx} \quad (4)$$

We finally obtain the probability P as single value indicating the quality of a registration result.

4. Results

For tests DR images acquired by Varian 4030R flat panels with 2304x3200 pixels resolution and DRRs with 512x512 resolution of two Alderson head phantoms and a pelvis phantom are used. The CT data for DRR reconstruction was recorded with 0.5 mm to 0.8 mm slice distance. Different types of errors were introduced, as misalignment between images that can be found by rigid registration, misalignment produced by 3D rotation of the imaged object and different degrees of salt and pepper noise. The probability was computed for a result lying within $TX = \pm0.5\ mm$, TY

$= \pm0.5~mm$ and $RZ = \pm0.5°$. Erroneous registrations that can be found by rigid registration could be identified in all cases. If a threshold at 40% probability is used for acceptance of the registration results, no false positives and negatives occur (Fig. 3a).
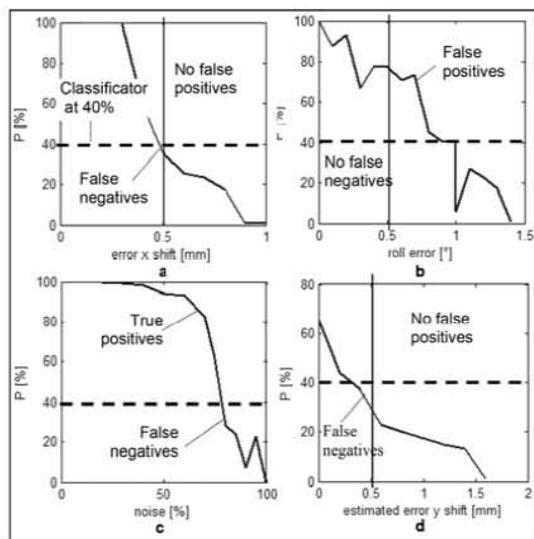

Figure 3. Probabilities a) for erroneous x-shifts; b) for rotations not detectable in 2D; c) for correctly registered images with salt and pepper noise; d) for images with significant radiometric differences

Fig. 3b shows the results for different rotations of the 3D object that cannot be found by 2D registration. Considering results below 40% probability as not accepted, we receive a certain amount of false positives because the images change only slightly with the 3D rotation (the rotation axis is parallel to the image plane). However, if we tolerate larger misalignments (e.g. 1.0°) we could also receive neither false positives nor false negatives.

We also tested the dependency on different noise levels in the floating image (Fig. 3c). The images were correctly registered and should always lie at a high probability. If results below 40% are rejected, we get false negatives not before 80% noise, which shows that the approach is very stable and still reliable at a high degree of noise.

Using images with significant radiometric variations coming from differences between the physical X-ray imaging and the DRR rendering, the results (see Fig. 3d) were comparable to those in Fig. 3a. In some cases when the images contained different internal structures, the large histogram variance led to low probabilities for the accepted misalignment of ±0.5 mm.

5. Conclusion

The proposed methods are suitable to provide a quality indicator for registration results in general. Up to a certain degree of noise or radiometric mismatch they are very stable and provide a straightforward possibility to classify registration results.

We suggest to use a Traffic-light type classification [5], e.g. at $P > 70\%$ (green - no false positives) and $P < 30\%$ (red - no false negatives). In our case, using the registration in the scope of a patient alignment system, we prefer to directly present the probability to the user. The system is operated by personal with a technical background and the quality indicator in percent is a meaningful and easy to interpret value.

References

[1] J. Pluim, J. Maintz, M. Viergever: Mutual information based registration of medical images. *IEEE transactions on medical imaging*, 22 (8): 986-1004, 2003.

[2] L. J. Verhey, M. Goitein, P. McNulty, J. E. Munzenrider, H. D. Suit: Precise positioning of patients for radiation therapy. *International Journal of Radiation Oncology Biol. Phys.*, 8 (2): 289-294, 1982.

[3] B. P. Selby, G. Sakas, S. Walter: 3D Alignment Correction for Proton Beam Treatment. *Biomedical Engineering*, 52: 2007.

[4] S. Clippe, D. Sarrut, C. Malet, S. Miguet, C. Ginestet, C. Carrie: Patient setup error measurement using 3D intensity-based image registration techniques. *International Journal of Radiation Oncology Biol. Phys.*, 56 (1): 259-265, 2003.

[5] W. Förstner: 10 pros and cons against performance characterization of vision algorithms. In: H. I. Christensen (eds.), *Workshop on Performance Characteristics of Vision Algorithms*, 9: 13-29, 1996.

[6] K. S. H. Tainter, U. Taneja, R. A. Robb: Quantitative validation of 3D image registration techniques. *Proceedings of SPIE Medical Imaging*, 2434: 504-519,1995.

[7] S. Hinz, A. Baumgartner: Urban Road Net Extraction Integrating Internal Evaluation Models. *Proceedings of the ISPRS-Commision III Symposium on Photogrammetric Computer Vision*, 34, 2002.

[8] S. Yu: Medical Image Registration Quality Analysis Using Neural Network Technique. Neurogenetics at UT Health Science Center, *http://www.nervenet.org*.

[9] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery: *Numerical Recipes in C 2*. Cambridge University Press, 1992.

[10] E. Parzen: On the estimation of a probability density function and the mode. Annals of Mathematical Statistics 33 (3): 1065-1076, 1962.