

Multi-Modal Registration of MR Images with a Novel Least-Squares Distance Measure

Stefan Heldmann^{a,b}

^aFraunhofer MEVIS, Institute for Medical Image Computing, Bremen, Germany

^bInstitute of Mathematics, University of Lübeck, Germany

ABSTRACT

In this work we evaluate a novel method for multi-modal image registration of MR images. The key feature of our approach is a new distance measure that allows for comparing modalities that are related by an arbitrary gray-value mapping. The novel measure is formulated as least square problem for minimizing the sum of squared differences of two images with respect to changing gray-values of one of the images. It turns out that the novel measure can be computed explicitly and allows for very simple and efficient implementation. We compare our new approach to rigid registration with cross-correlation, mutual information, and normalized gradient fields as distance measure.

1. INTRODUCTION

The main ingredient for any intensity based image registration method is a distance measure that quantifies similarity of images. State-of-the-art approaches paying tribute to multi-modality are mainly based on correlation, mutual information, or edges of images.¹⁻⁵ These type of measure are able to compare images that stem from different modalities since they are invariant with respect to certain classes of gray-value transformations. For example, cross-correlation^{1,2,5} (CC) is invariant with respect to affine linear gray-value changes which is achieved by implicit normalization of images with their mean and standard deviation. That is, if we consider two images that are identical up to an affine linear transformation of gray-values then the cross-correlation measure will classify them as identical. More involved measures such as the correlation ratio (CR),⁴ mutual information^{3,5-7} (MI), or normalized gradient fields⁸ (NGF) are much more powerful. They are invariant with respect to arbitrary gray-value transformations and in the case of MI and NGF the transformation can be even spatially dependent. In the case of mutual information this is achieved through comparing statistics of the joint gray-value densities of images. The NGF distance measure considers edge information which also is independent of particular intensities.

However, all these approaches have different advantages and drawbacks. The cross-correlation measure is easy to compute but it is only meaningful if two modalities are related by a mapping that is close to being globally affine linear. The correlation ratio is much more general but it is defined as a rational expression which makes computing derivatives involved and complicates its use in a registration algorithm that relies on derivative based optimization. Mutual information is very powerful and allows for a meaningful comparison of arbitrary modalities. Drawbacks are that it is not easy to interpret, an implementation is involved and several variants have been proposed in literature leading to implementation depending results. The NGF measure has been introduced to overcome latter drawbacks. The key of this measure is that it considers derivatives of images providing edge information. Therefore, one needs to estimate image gradients that are sensitive to noise in the images.

In our new approach we use a novel measure that can be seen as a generalization of the sum-of-squared differences measure (SSD), which per se is only suitable for images of same modality, to the multi-modal case. The new measure is initially defined as a least-square problem by minimizing the SSD of two images with respect to arbitrary gray-value transformations.

It turns out, the new measure is nothing else than a weighted 2-norm that is easy to implement and allows for efficient computation. In essence, the new measure classifies two images identical if and only if they differ with respect to a global transformation of intensities. Therefore, it is similar to the correlation ration, much more powerful than cross correlation that is invariant only with respect to affine linear dependency of intensities, and

bit less powerful than, e.g., mutual information which also provides meaningful comparisons for images related by spatially dependent gray-value transformations.

In this work we demonstrate the effectiveness of our new measure for rigid registration of different types of MR images and compare them to the results obtained for cross-correlation, mutual information, and normalized gradient fields as distance measures.

The paper is organized as follows. We start with setting the scene for intensity based rigid registration and introduce the novel distance measure. Subsequently, we show how it can be computed efficiently and give some details of our registration algorithm based on Gauss-Newton optimization. Finally we present experimental results for multi-modal registration of MR images and a discuss our work.

2. METHODS

The general task of image registration is to compute a spatial alignment of images. Here, we assume we are given two images R and T , where R is called reference and T is called template. The main building block of any registration method is a criteria that quantifies the quality of the aligned images, i.e., their similarity. We measure the similarity of images by a so-called distance measure \mathcal{D} such that $\mathcal{D}(R, T)$ is small if R and T are similar and $\mathcal{D}(R, T)$ is large if not. For sake of simplicity we consider discrete images R and T made up from totally N pixels/voxels and furthermore that the images are reshaped as N -vectors such that $R, T \in \mathbb{R}^N$. Therefore, the distance measure is a mapping $\mathcal{D} : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$. The goal of the registration is then to find a deformation, such that the reference and the deformed template becomes similar. On that account, we perform registration by minimizing the distance of reference and template with respect to the deformation. In this work we particularly consider rigid deformations, i.e., we only allow rotation and translation of the template yielding 3 degrees of freedom in 2D (rotation and translation in x -, y - direction) and 6 degrees of freedom for three-dimensional images (rotation around x -, y -, z -axis, translation in x -, y -, z -direction). Let θ be a parameter vector collecting the transformation parameters for rotation and translation and let $T(\theta) \in \mathbb{R}^N$ denote the rigidly deformed template image. The rigid registration is then performed by computing optimal parameters θ such that

$$\mathcal{D}(R, T(\theta)) \stackrel{!}{=} \min. \quad (1)$$

We will present the details of our algorithm in section 2.3. The novelty of our approach is a new distance measure for multi-modal registration, which we describe next.

2.1 The Least Squares Distance Measure

The problem when measuring similarity of images from different modalities is that intensities cannot be compared directly. That is, the same object will be mapped to different gray-values depending on the particular modality. To overcome this problem we present a novel measure that is invariant with respect to arbitrary gray-value transformations. To make the idea clear, consider displaying an image. Then we can create multiple views of the same image by choosing different colormaps. The same principle applies to our new measure. If reference and template are identical up to a particular gray-value labeling, such that $T = g(R)$ where $g : \mathbb{R} \rightarrow \mathbb{R}$ is an arbitrary gray-value transform, then it will classifies them as identical, i.e., it takes its minimum.

The starting point for deriving our new measure is the popular sum-of-squared differences measure SSD. It is defined as

$$\text{SSD}(R, T) := \frac{1}{2} \|T - R\|^2 = \frac{1}{2} \sum_{j=1}^N (T_j - R_j)^2.$$

The SSD is easy to compute and easy to interpret but it clearly relies on images with comparable gray-values for meaningful comparison. In particular $\text{SSD}(R, T)$ will take its minimum (which is zero) if and only if R and T are identical, that is

$$\text{SSD}(R, T) = 0 \quad \Leftrightarrow \quad R = T.$$

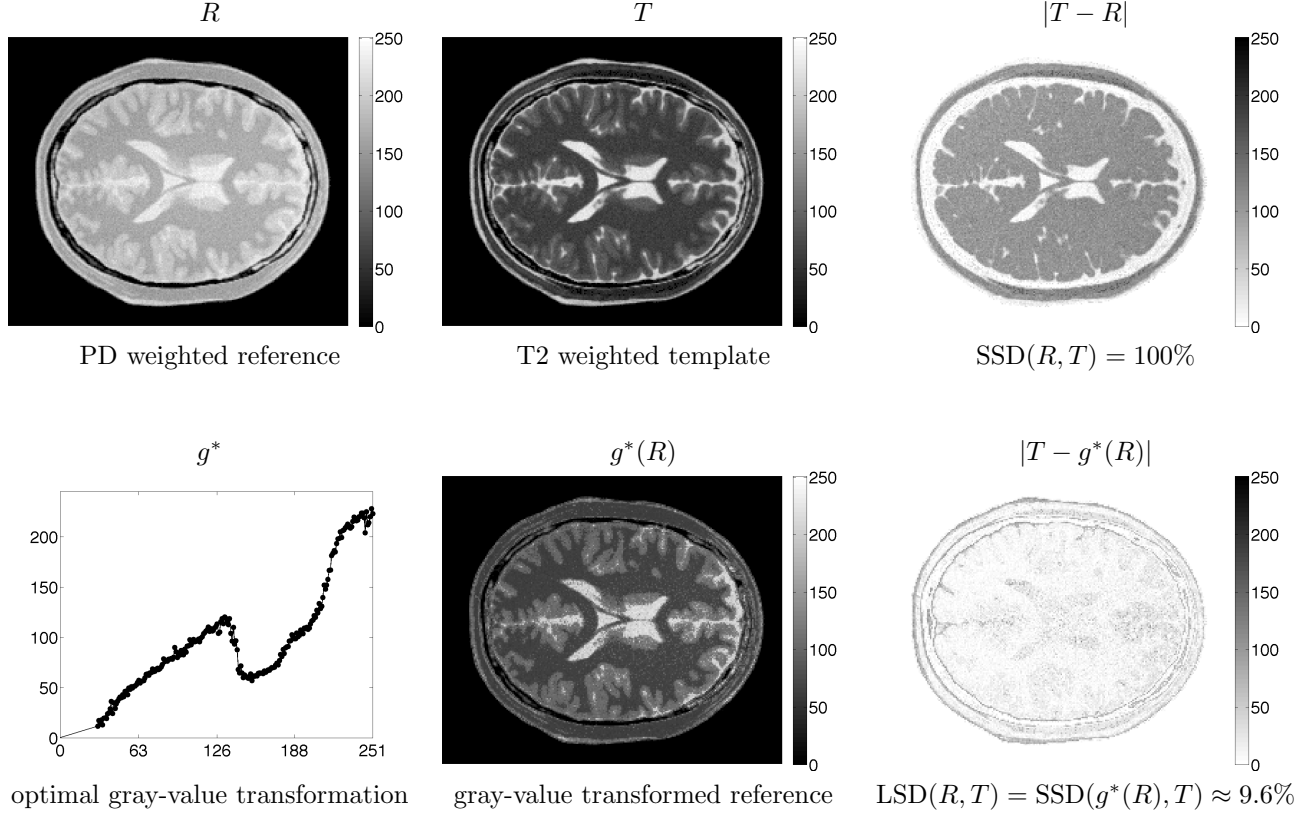


Figure 1. Measuring the least-square distance LSD of two simulated MR brain images. Top row from left to right: proton density weighted MR reference image R ; T2-weighted MR template image T ; difference between reference and template. Bottom row from left to right: optimal gray-value transformation g^* for the reference (x -axis = gray-value range of R , y -axis = gray-value range of T); optimal gray-value transformed reference $g^*(R)$; difference between optimal gray-value transformed reference and template

Thus, it is in general only suitable for mono-modal registration. For our new measure, we want that it is invariant with respect to changing gray-values of an image. For that reason, we define our new least-squares-distance measure (LSD) as the least-squares problem

$$\text{LSD}(R, T) := \inf_{g: \mathbb{R} \rightarrow \mathbb{R}} \frac{1}{2} \|T - g(R)\|^2 = \inf_{g: \mathbb{R} \rightarrow \mathbb{R}} \frac{1}{2} \sum_{j=1}^N (T_j - g(R_j))^2, \quad (2)$$

where $g(R)$ is meant point-wise, i.e., $g(R) = (g(R_1), g(R_2), \dots, g(R_N))^T$. Thus, it takes its minimum if and only if R and T are identical up to gray-value changes, i.e.,

$$\text{LSD}(R, T) = 0 \Leftrightarrow \text{there exists } g: \mathbb{R} \rightarrow \mathbb{R} \text{ such that } g(R) = T.$$

Note that the least-squares distance can also be expressed in terms of the sum of squared differences, which justifies viewing it as an extension of the SSD measure, since

$$\text{LSD}(R, T) = \inf_{g: \mathbb{R} \rightarrow \mathbb{R}} \text{SSD}(g(R), T).$$

On first sight, the new distance measure looks involved, since it requires solving a least squares problem. In the following, we show that the optimal gray-value transformation g can be computed explicitly yielding a concise formulation of the LSD measure which is easy to compute and allows for very efficient implementation.

Figure 1 illustrates the idea of the least-squares distance. Here we compare two simulated MR brain images taken from the Brainweb data base,^{9,10} where both images are identical up to their (simulated) modality. Then we compute the optimal gray-value transformation g^* and apply it to the reference. As a result, the gray-value transformed reference $g^*(R)$ is close to an image with same modality as the template. Therefore, we can compare T and $g^*(R)$ directly and measure their distance with a mono-modal measure such as SSD. In particular, the least squares distance is given by $\text{LSD}(R, T) = \text{SSD}(g^*(R), T)$. Defining the value of the SSD of R and T as 100% we find that their least-squares distance is only approximately 9.6% which means the images are very similar.

Clearly, the LSD measure is not symmetric. That is, in general comparing T to R gives a different value than comparing R to T and in general holds $\text{LSD}(R, T) \neq \text{LSD}(T, R)$. However, since we are interested in using the LSD as a distance measure for image registration and in general image registration by itself is not symmetric unless explicitly modeled,¹¹ we do not see this as a major drawback. In our experiment we considered all combination of images, such that each image is used as a reference as well as template. With respect to our experiments, the results show, that the particular choice of an image as reference or template was not crucial for the outcomes of the registrations and yield qualitatively the same results.

2.2 Computing the Least-Squares Distance Measure

With help of some linear algebra we are able to compute a solution to the least-square problem and to establish another surprisingly simple formulation of the LSD measure. In the following we are mainly interested in the mapping $T \mapsto \text{LSD}(R, T)$ with a fixed reference image R as motivated by the registration problem (1). For that reason and for simple presentation, we omit any notational overhead indicating the dependence of a quantity on the reference.

Assume the reference R takes only gray values in the set $\{r_1, \dots, r_M\}$ with $1 \leq M \leq N$ such that for each gray value r_k there exists at least one index j with $R_j = r_k$. Next, we define the matrix

$$Q \in \{0, 1\}^{N \times M} \quad \text{with} \quad Q_{jk} = \begin{cases} 1 & \text{if } R_j = r_k, \\ 0 & \text{else} \end{cases} \quad (3)$$

and, with some abuse of notation, we collect the values $g(r_k)$ of the gray value transform in a vector we also denote by g , i.e., we set $g \equiv (g(r_1), \dots, g(r_M))^T$. Then we can rewrite the image R as the matrix-vector product $R = Qg$. For better understanding, here is a small example of decomposing a 3×3 image made up from three gray-values $r_1 = 10$, $r_2 = 5$ and $r_3 = 2$ into the product Qg :

$$R = \begin{array}{|c|c|c|} \hline 10 & 5 & 10 \\ \hline 5 & 2 & 5 \\ \hline 10 & 5 & 10 \\ \hline \end{array} \quad \rightsquigarrow \quad R = \begin{pmatrix} 10 \\ 5 \\ 10 \\ 5 \\ 2 \\ 5 \\ 10 \\ 5 \\ 10 \end{pmatrix} = 10 \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + 5 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}}_{=Q} \underbrace{\begin{pmatrix} 10 \\ 5 \\ 2 \end{pmatrix}}_{=g} = Qg$$

The matrix Q is sparse with exactly N non-zeros entries. Furthermore, the columns of Q are pairwise orthogonal (but not orthonormal) with at least one non-zero entry such that Q has full rank M . The last property holds because we required that for each gray value r_k exists at least one index j with $R_j = r_k$. Thus, we obtain

$$\frac{1}{2} \|T - g(R)\|^2 = \frac{1}{2} \|T - Qg\|^2 \quad (4)$$

and the optimal gray value transform g^* minimizing (4) is given by

$$g^* = (Q^T Q)^{-1} Q^T T. \quad (5)$$

Note that $Q^\top Q$ is a M -by- M matrix such that the full rank M of Q ensures that $Q^\top Q$ has also full rank and therefore is invertible. Furthermore, as mentioned above, the columns of Q are pairwise orthogonal, such that $Q^\top Q$ is a diagonal matrix and therefore its inverse $(Q^\top Q)^{-1}$ is diagonal, too. Further inspection shows, that the entries on the diagonal are the values of the histogram of R . For the k -th diagonal entry $(Q^\top Q)_{kk}$ we find

$$(Q^\top Q)_{kk} = e_k^\top Q^\top Q e_k = \sum_{j=1}^N Q_{jk}^2 = \#\{j : R_j = r_k\} =: \text{hist}_k(R)$$

where e_k denotes the k -th unit vector of \mathbb{R}^N . Thus, $Q^\top Q$ and its inverse are given by

$$Q^\top Q = \begin{pmatrix} \text{hist}_1(R) & & \\ & \ddots & \\ & & \text{hist}_M(R) \end{pmatrix} \quad \text{and} \quad (Q^\top Q)^{-1} = \begin{pmatrix} 1/\text{hist}_1(R) & & \\ & \ddots & \\ & & 1/\text{hist}_M(R) \end{pmatrix}. \quad (6)$$

Hence, we can directly compute the optimal gray-value transformation g^* rather than solving a linear system. However, combining (4) and (5) we find that the least-square distance (2) is given by

$$\text{LSD}(R, T) = \inf_{g: \mathbb{R} \rightarrow \mathbb{R}} \frac{1}{2} \|T - g(R)\|^2 = \frac{1}{2} \|T - Qg^*\|^2 = \frac{1}{2} \|PT\|^2 \quad (7)$$

with

$$P := I - Q(Q^\top Q)^{-1}Q^\top. \quad (8)$$

Thus, we have found an explicit formulation of the least-square distance. The structure of the matrix P allows for one step further simplification of (7). Rewriting the 2-norm as inner product we obtain $\|PT\|^2 = T^\top P^\top PT$ and moreover

$$\begin{aligned} P^\top P &= (I - Q(Q^\top Q)^{-1}Q^\top)^\top (I - Q(Q^\top Q)^{-1}Q^\top) \\ &= I - Q(Q^\top Q)^{-1}Q^\top - (Q(Q^\top Q)^{-1}Q^\top)^\top + (Q(Q^\top Q)^{-1}Q^\top)^\top Q(Q^\top Q)^{-1}Q^\top \\ &= I - Q(Q^\top Q)^{-1}Q^\top \\ &= P. \end{aligned} \quad (9)$$

Hence, $\|PT\|^2 = T^\top P^\top PT = T^\top PT$ and the LSD can be computed as

$$\text{LSD}(R, T) = \frac{1}{2} T^\top PT.$$

Note that (9) implies that P is a symmetric positive semi definite matrix.

Summarizing, we have found an explicit formulation of the LSD measure, that is easy to evaluate and allows for efficient implementation. Since the matrix Q is sparse with exactly N nonzero entries and the inverse $(Q^\top Q)^{-1}$ is a M -by- M diagonal matrix with $M < N$, we can effectively compute the matrix-vector product PT with $\mathcal{O}(N)$ arithmetic operations. Thus, the computational complexity for computing the LSD measure is $\mathcal{O}(N)$.

2.3 Rigid Registration

As in the beginning of section 2, let $\theta \in \mathbb{R}^p$ be a vector that holds the rotation and translation parameters of a rigid transformation ($p = 3$ in 2D and $p = 6$ in 3D). Furthermore, let $T(\theta) \in \mathbb{R}^N$ denote the rigidly deformed template image, i.e., we consider the template image as a mapping $T : \mathbb{R}^p \rightarrow \mathbb{R}^N$ with $\theta \mapsto T(\theta)$. For the rigid registration we then minimize the objective function

$$F(\theta) := \text{LSD}(R, T(\theta)) = \frac{1}{2} T(\theta)^\top PT(\theta) \quad (10)$$

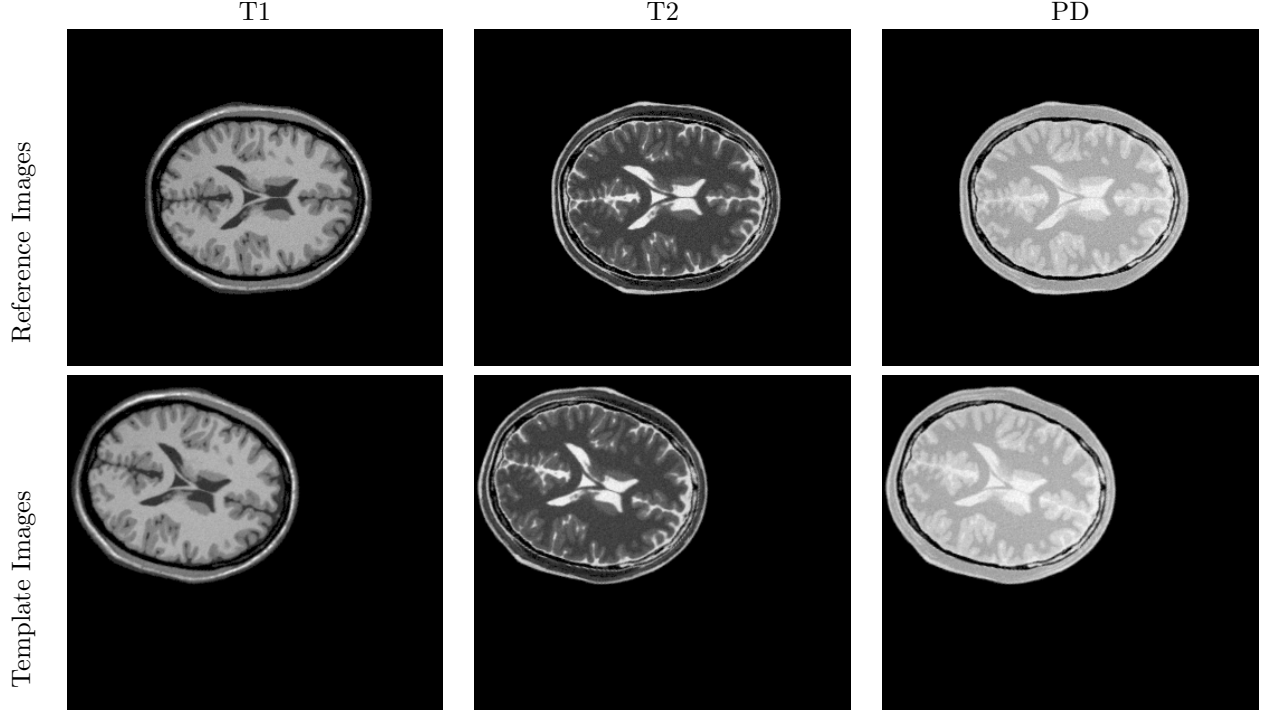


Figure 2. Reference and template images used in the experiments.

with matrix the P defined for the fixed reference image R , cf. (3) and (8). For the minimization of (10) we use a Gauss-Newton method.¹² Therefore, we need to compute the gradient

$$\nabla F(\theta) = \nabla T(\theta)^\top P T(\theta) \in \mathbb{R}^p,$$

where $\nabla T(\theta)$ is the N -by- p Jacobian matrix of T . Furthermore we use and a Gauss-Newton approximation to the Hessian $\nabla^2 F$ of F by neglecting terms involving second order derivatives of the template image. Here we use the approximation

$$\nabla T(\theta)^\top P \nabla T(\theta) \approx \nabla^2 F(\theta) \in \mathbb{R}^{p \times p}.$$

For the registration we then performed the following iteration:

Choose initial guess θ_0

for $k = 0, 1, 2, \dots$ do

 Compute gradient $g_k := \nabla F(\theta_k)$ and approximate Hessian $H_k := \nabla T(\theta_k)^\top P \nabla T(\theta_k)$

 Compute search direction $s_k = -H_k^{-1} g_k$

 Compute step-length $\alpha > 0$ and update $\theta_{k+1} = \theta_k + \alpha s_k$

 if any stopping criteria is fulfilled then STOP

end

3. RESULTS

We test our new method with artificially deformed T1, T2, and proton density weighted MR images taken from the brainweb data-base (www.bic.mni.mcgill.ca/brainweb). The template images were generated from the reference images by rotating 20 degree around the center and subsequent vertical and horizontal translation by 30 and 60 pixels. The test images are shown in Figure 2. We compare the registration with our new method with

MODALITIES ($R-T$)	ROTATION ERRORS (DEGREE)			
	CC	LSD	MI	NGF
T1-T1	$4.33 \cdot 10^{-06}$	$1.97 \cdot 10^{-14}$	$4.17 \cdot 10^{-06}$	$1.59 \cdot 10^{-06}$
T1-T2	$3.49 \cdot 10^{+01}$	$3.74 \cdot 10^{-02}$	$4.08 \cdot 10^{-03}$	$2.59 \cdot 10^{-01}$
T1-PD	$3.98 \cdot 10^{+00}$	$8.90 \cdot 10^{-06}$	$3.61 \cdot 10^{-01}$	$3.38 \cdot 10^{-01}$
T2-T1	$6.84 \cdot 10^{+01}$	$4.39 \cdot 10^{-01}$	$3.40 \cdot 10^{-01}$	$2.07 \cdot 10^{-01}$
T2-T2	$5.52 \cdot 10^{-06}$	$1.84 \cdot 10^{-14}$	$7.55 \cdot 10^{-06}$	$4.03 \cdot 10^{-06}$
T2-PD	$3.63 \cdot 10^{-06}$	$4.93 \cdot 10^{-02}$	$1.77 \cdot 10^{-01}$	$2.23 \cdot 10^{-01}$
PD-T1	$3.87 \cdot 10^{+00}$	$2.26 \cdot 10^{-01}$	$3.94 \cdot 10^{-01}$	$2.10 \cdot 10^{-01}$
PD-T2	$9.14 \cdot 10^{-03}$	$4.07 \cdot 10^{-01}$	$1.21 \cdot 10^{-02}$	$1.52 \cdot 10^{-01}$
PD-PD	$8.64 \cdot 10^{-06}$	$4.74 \cdot 10^{-08}$	$3.88 \cdot 10^{-06}$	$2.27 \cdot 10^{-06}$

MODALITIES ($R-T$)	TRANSLATION ERRORS (PIXEL)			
	CC	LSD	MI	NGF
T1-T1	$5.28 \cdot 10^{-06}$	$2.11 \cdot 10^{-14}$	$1.22 \cdot 10^{-06}$	$4.88 \cdot 10^{-06}$
T1-T2	$3.47 \cdot 10^{+00}$	$3.92 \cdot 10^{-02}$	$1.55 \cdot 10^{-01}$	$3.63 \cdot 10^{-02}$
T1-PD	$3.45 \cdot 10^{-01}$	$1.33 \cdot 10^{-02}$	$5.89 \cdot 10^{-02}$	$1.11 \cdot 10^{-01}$
T2-T1	$4.54 \cdot 10^{-01}$	$5.57 \cdot 10^{-02}$	$1.53 \cdot 10^{-02}$	$3.82 \cdot 10^{-02}$
T2-T2	$3.98 \cdot 10^{-06}$	$1.56 \cdot 10^{-14}$	$1.22 \cdot 10^{-05}$	$5.92 \cdot 10^{-06}$
T2-PD	$3.20 \cdot 10^{-06}$	$1.87 \cdot 10^{-01}$	$2.42 \cdot 10^{-02}$	$4.77 \cdot 10^{-02}$
PD-T1	$3.65 \cdot 10^{-01}$	$6.72 \cdot 10^{-02}$	$5.09 \cdot 10^{-02}$	$1.33 \cdot 10^{-02}$
PD-T2	$3.85 \cdot 10^{-01}$	$6.76 \cdot 10^{-02}$	$1.46 \cdot 10^{-01}$	$5.40 \cdot 10^{-03}$
PD-PD	$7.10 \cdot 10^{-06}$	$1.95 \cdot 10^{-08}$	$5.52 \cdot 10^{-06}$	$1.84 \cdot 10^{-05}$

Table 1. Registration Errors. Rotation errors larger than 1 degree and translation errors larger than 1 pixel are highlighted.

rigid registration using cross-correlation, mutual information and normalized gradients as distance measures. In particular for the registration based on mutual information we used a BFGS quasi Newton method^{12,13} and B-spline kernel density estimation.¹³ Furthermore, to avoid local minima, for all tests we used a multi-level scheme. Therefore we computed an image pyramid by subsequent smoothing and down-sampling the reference and template images. Starting from the coarsest level with coarsest resolution we performed a registration and used the results as initial guess for the next higher level with higher resolution until we reach the finest level with the original images.

The resulting errors between artificial and recovered transformation parameters for rotation and translation are presented in Table 1. The results show that the new method using the LSD measure performs comparable to the registration based on mutual information and normalized gradient fields. Furthermore, the registration based on cross-correlation worked basically only in mono-modal cases and failed for T1,T2 weighted images and T1,PD weighted images.

4. CONCLUSION

We successfully evaluated our new method in an experimental setting for multi-modal registration of MR images. We found that our new approach worked competitive compared to well established multi-modal methods based on mutual information and normalized gradient fields. Furthermore, the experiments showed that our approach outperformed registration using cross-correlation as a distance measure. Compared to the other methods, the main advantages of our approach are that it is easy to understand and to implement. With respect to the experiments done, we conclude that LSD is a simple alternative to MI and NGF.

REFERENCES

1. L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys* **24**(4), pp. 325–376, 1992.
2. R. P. Woods, S. R. Cherry, and J. C. Mazziotta, "Rapid automated algorithm for aligning and reslicing pet images," *J Comput Assist Tomogr* **16**(4), pp. 620–33, 1992.
3. P. A. Viola and W. M. I. Wells, "Alignment by maximization fo mutual information," in *5th International Conference on Computer Vision*, 1995.
4. A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *MICCAI 1998, LNCS 1496, Lecture Notes in Computer Science 1496*, pp. 1115–1124, Springer Verlag, (Cambridge, MA), Oct. 1998.
5. I. N. Bankman, ed., *Handbook of medical imaging*, Academic Press, Inc., Orlando, FL, USA, 2000.
6. A. Collignon, F. Maes, P. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registartion based on information theory," in *Computational Imaging and Vision: Proceedings XIVth international conference on information processing in medical imaging - IPMI'95*, I. d. Berder, B. Y. France, B. C., and R. Di Paola, eds., *Information Processing in Medical Imaging 3*, pp. 263–274, Kluwer Academic Publishers, 1995.
7. J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual information based registration of medical images: a survey," *IEEE Transactions on Medical Imaging* **22**, pp. 986–1004, 2003.
8. E. Haber and J. Modersitzki, "Intensity gradient based registration and fusion of multi-modal images," *Methods of Information in Medicine* **46**(3), pp. 292–299, 2007.
9. C. A. Cocosco, V. Kollokian, R. K. S. Kwan, G. B. Pike, and A. C. Evans, "Brainweb: Online interface to a 3d mri simulated brain database," *NeuroImage* **5**, 1997.
10. "Brainweb: Simulated brain database." <http://www.bic.mni.mcgill.ca/brainweb>.
11. G. E. Christensen and J. He, "Consistent nonlinear elastic image registration," in *MMBIA '01: Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '01)*, p. 37, IEEE Computer Society, (Washington, DC, USA), 2001.
12. J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer, 1999.
13. D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Transactions on Medical Imaging* **22**(1), pp. 120–128, 2003.