

MULTI-FEATURE MUTUAL INFORMATION IMAGE REGISTRATION

DEJAN TOMAŽEVIČ^{✉,1,2}, BOŠTJAN LIKAR¹ AND FRANJO PERNUŠ¹

¹University of Ljubljana, Faculty of Electrical Engineering Tržaška 25, SI-1000 Ljubljana, Slovenia; ²Sensum, Computer Vision Systems, Tehnološki park 21, SI-1000 Ljubljana, Slovenia
e-mail: dejan.tomazevic@fe.uni-lj.si, bostjan.likar@fe.uni-lj.si, franjo.pernus@fe.uni-lj.si
(Received September 5, 2011; revised December 18, 2011; accepted December 23, 2011)

ABSTRACT

Nowadays, information-theoretic similarity measures, especially the mutual information and its derivatives, are one of the most frequently used measures of global intensity feature correspondence in image registration. Because the traditional mutual information similarity measure ignores the dependency of intensity values of neighboring image elements, registration based on mutual information is not robust in cases of low global intensity correspondence. Robustness can be improved by adding spatial information in the form of local intensity changes to the global intensity correspondence. This paper presents a novel method, by which intensities, together with spatial information, *i.e.*, relations between neighboring image elements in the form of intensity gradients, are included in information-theoretic similarity measures. In contrast to a number of heuristic methods that include additional features into the generic mutual information measure, the proposed method strictly follows information theory under certain assumptions on feature probability distribution. The novel approach solves the problem of efficient estimation of multi-feature mutual information from sparse high-dimensional feature space. The proposed measure was tested on magnetic resonance (MR) and computed tomography (CT) images. In addition, the measure was tested on positron emission tomography (PET) and MR images from the widely used Retrospective Image Registration Evaluation project image database. The results indicate that multi-feature mutual information, which combines image intensities and intensity gradients, is more robust than the standard single-feature intensity based mutual information, especially in cases of low global intensity correspondences, such as in PET/MR images or significant intensity inhomogeneity.

Keywords: computed tomography, magnetic resonance, multi-feature mutual information, positron emission tomography, registration, similarity measure

INTRODUCTION

Geometric alignment or registration of images is a fundamental task in numerous computer-assisted applications in medical imaging (Hawkes, 1998; Maintz and Viergever, 1998). Finding spatial relations between two or more images of the same modality and of the same subject is useful for observing and quantifying changes that anatomical structures undergo during a period of time, whereas registration of multimodal images is important for combining the complementary information of different imaging modalities. Furthermore, registration of images of different subjects allows comparative studies, which are needed to capture how a structure and its function vary in large populations, across age and gender and in different disease states. Registration algorithms can be classified according to the image features and the feature correspondence criterion used to compute geometric transformations between images. According to image features, registration algorithms can be divided into:

fiducial-based, segmentation-based, and image-based. Image-based registration methods optimize a similarity measure between image features, which are usually intensities of image elements, *i.e.*, pixels or voxels. The comparison and evaluation of retrospective inter-modality brain registration techniques has shown that the best registration results were obtained by the image-based registration technique using mutual information as the similarity measure (West *et al.*, 1997).

Mutual information was first proposed for the purpose of image registration by Collignon *et al.* (1995) and Viola and Wells (1995), independently. The idea was further developed in their later publications (Wells *et al.*, 1996; Maes *et al.*, 1997). Since then, the mutual information measure, as well as its derivatives, have become the most popular and most studied similarity measure in the medical image registration community. The original idea of Collignon *et al.* (1995) and Viola and Wells (1995) has been implemented and modified by various researchers in their registration studies, resulting in numerous publications, a survey

of which can be found in Pluim *et al.* (2003) and Maes *et al.* (2003). The main reason for the frequent use of the mutual information measure is that it only assumes that image features have a probabilistic relationship which is the loosest of assumptions which can be made. However, mutual information based registration can fail when statistical dependence between image features is weak or as a result of interpolation artifacts. Weak dependence between image intensities can either be due to fundamentally different imaging modalities, such as MR and ultrasound (Roche *et al.*, 2000), noise (Tsao, 2003), low resolution of images (Pluim *et al.*, 2000b) or small image overlap (Studholme *et al.*, 1999). Interpolation artifacts are often enhanced due to image noise and low resolution of images (Pluim *et al.*, 2000b; Tsao, 2003), but can be reduced by higher order interpolation schemes (Thevenaz and Unser, 2000). Because the traditional mutual information similarity measure ignores the dependency of feature values of neighboring image elements, some researchers argued that the problem of weak statistical dependence of intensities could be resolved by adding explicit spatial information into the registration process (Pluim *et al.*, 2000a; Rueckert *et al.*, 2000). Rodríguez-Carranza and Loew (2000) proposed to use the Jumarie S-entropy, which considers the difference of gray values of neighboring image elements. Pluim *et al.* (2000a) incorporated spatial information by multiplying the mutual information similarity measure with the similarity measure of gradients of corresponding points. Their results show that without the loss of accuracy the combined criterion is more robust than standard mutual information. However, ad hoc multiplication of two or more different criteria opens the problem of controlling the amount of contribution of each criterion. Rueckert *et al.* (2000) introduced an extension of mutual information called second order mutual information, which was estimated from the co-occurrence of gray values of neighboring image elements within the image and gray values of corresponding image elements of images to be registered. The approach requires estimation of joint probability distributions from a four-dimensional joint feature histogram. A potential disadvantage of this approach is that the high dimensionality of the joint histogram decreases the power of histogram statistics and makes estimation of the similarity measure less reliable. Estimation of high dimensional probability distributions can be avoided if entropy can be estimated directly from feature samples. In the work of Hero *et al.* (2002) it was shown that joint entropy can be efficiently estimated from the length of minimum spanning trees (MST) that connect all samples in the feature space. For multi-feature registration, the MST approach was

implemented in the work of Sabuncu and Ramadge (2003), Neemuchwala *et al.* (2005), Neemuchwala *et al.* (2006) and Staring *et al.* (2009). A potential disadvantage of MST registration as introduced by Hero *et al.* (2002) is the time-consuming optimization that is required to obtain a minimum spanning tree for each evaluation of similarity measure. Sabuncu and Ramadge (2008) have shown how the alignment measure and a descent direction with respect to alignment parameters can be determined directly from MST, which makes MST-based registration more efficient. A different approach to more effective joint entropy estimation has been introduced by Špiclin *et al.* (2012), where the authors propose gradient based joint entropy minimization by implementation of Hilbert kernel joint density estimation.

In this paper we propose an extension of the standard single-feature mutual information similarity measure to a multi-feature mutual information measure, and solve the problem of efficient estimation of feature joint probability distribution in a high-dimensional feature space. The proposed measure incorporates spatial information by combining the commonly used image intensity feature with additional features. Additional features, such as intensity gradients or second order derivatives may be derived from basic features or obtained by multi-channel imaging. Furthermore, the proposed multi-feature mutual information measure can be computed for an arbitrary number of additional features, as long as the probability distributions of these features are approximately normal.

MATERIALS AND METHODS

Image registration is concerned with finding the geometrical transformation T that brings a floating image A into the best possible spatial correspondence with a reference image B . Let images A and B be represented by image features $z_a(x)$ and $z_b(x)$, respectively, where $z(x)$ is the value of image feature z at position x in image space. Let SM denote the criterion function or similarity measure between two corresponding feature sets $z_a(x)$ and $z_b(T(x))$ for a given transformation T . Registration is then concerned with finding the spatial transformation \hat{T} that maximizes the similarity measure SM

$$\hat{T} = \arg \max_T SM(z_a(x), z_b(T(x))). \quad (1)$$

The mutual information (MI) similarity measure assumes that image features $z_a(x)$ and $z_b(x)$ are observed values of random variables Z_a and Z_b , respectively. In terms of entropy, MI is defined as (Cover and Thomas, 1991)

$$MI(Z_a, Z_b) = H(Z_a) + H(Z_b) - H(Z_a, Z_b), \quad (2)$$

where $H(Z_a)$ and $H(Z_b)$ are corresponding entropies of random variables Z_a and Z_b . The entropy $H(Z)$ of a random variable Z (Z_a or Z_b) is defined as

$$H(Z) = -\int p(z) \log p(z) dz, \quad (3)$$

where $p(z)$ ($p(z_a)$ or $p(z_b)$) is the probability density function describing the probability of feature z (z_a or z_b). The third term in (Eq. 2), $H(Z_a, Z_b)$, is the joint entropy of random variables Z_a and Z_b and is defined as

$$H(Z_a, Z_b) = -\int \int p(z_a, z_b) \log p(z_a, z_b) dz_a dz_b, \quad (4)$$

where $p(z_a, z_b)$ is the joint probability density function, defining the probability of feature values z_a and z_b appearing at corresponding positions x and $T(x)$ in images A and B , respectively. In case of discrete values of image features, calculation of the entropy and the joint entropy can be simplified by calculating Shannon's entropy of discrete random variables

$$H(Z) = -\sum_{z_a} p(z) \log p(z). \quad (5)$$

In general, entropy is a measure of the amount of uncertainty, variability, or complexity of a random variable, whereas mutual information of two random variables is a measure of the amount of uncertainty of one random variable, knowing the value of the other variable.

NEW MULTI-FEATURE MUTUAL INFORMATION MEASURE

Let us extend the image representation from a single-feature $z(x)$ to a multi-feature representation $\mathbf{z}(x)$, where $\mathbf{z}(x)$ is a vector function $\mathbf{z}(x) = (z_1(x), \dots, z_K(x))$ defining the value of features $z_k(x)$, $k = 1 \dots K$, at position x in image space. The values of features $z_k(x)$ are observed values of random variable Z_k , comprising the vector of random variables \mathbf{Z} . Multi-feature mutual information of corresponding vectors of random variables \mathbf{Z}_a and \mathbf{Z}_b is defined as

$$\begin{aligned} MI(\mathbf{Z}_a, \mathbf{Z}_b) &= H(I_a, \mathbf{V}_a) + H(I_b, \mathbf{V}_b) - H(I_a, I_b, \mathbf{V}_a, \mathbf{V}_b) = \\ &= H(I_a) + H(\mathbf{V}_a | I_a) + H(I_b) + H(\mathbf{V}_b | I_b) - H(I_a, I_b) - H(\mathbf{V}_a, \mathbf{V}_b | I_a, I_b) = \\ &= H(I_a) + H(I_b) - H(I_a, I_b) + H(\mathbf{V}_a | I_a) + H(\mathbf{V}_b | I_b) - H(\mathbf{V}_a, \mathbf{V}_b | I_a, I_b) = \\ &= MI(I_a, I_b) + AI(I_a, I_b, \mathbf{V}_a, \mathbf{V}_b) \end{aligned} \quad (10)$$

$$AI(I_a, I_b, \mathbf{V}_a, \mathbf{V}_b) = H(\mathbf{V}_a | I_a) + H(\mathbf{V}_b | I_b) - H(\mathbf{V}_a, \mathbf{V}_b | I_a, I_b), \quad (11)$$

$$MI(\mathbf{Z}_a, \mathbf{Z}_b) = H(\mathbf{Z}_a) + H(\mathbf{Z}_b) - H(\mathbf{Z}_a, \mathbf{Z}_b), \quad (6)$$

where $H(\mathbf{Z}_a)$, $H(\mathbf{Z}_b)$ and $H(\mathbf{Z}_a, \mathbf{Z}_b)$ are the joint entropies of vectors of random variables $\mathbf{Z}_a = (Z_{a1}, \dots, Z_{aK})$, $\mathbf{Z}_b = (Z_{b1}, \dots, Z_{bK})$ and $(\mathbf{Z}_a, \mathbf{Z}_b) = (Z_{a1}, \dots, Z_{aK}, Z_{b1}, \dots, Z_{bK})$, respectively. In general, the entropy of a vector of random variables $\mathbf{Z} = (Z_1, \dots, Z_K)$ is defined as

$$\begin{aligned} H(\mathbf{Z}) &= -\int p(\mathbf{z}) \log p(\mathbf{z}) d\mathbf{z} = \\ &= -\int \dots \int p(z_1, \dots, z_K) \log p(z_1, \dots, z_K) dz_1 \dots dz_K \end{aligned}, \quad (7)$$

while in case when (Z_1, \dots, Z_K) are discrete random variables, the entropy $H(\mathbf{Z})$ may be obtained as

$$\begin{aligned} H(\mathbf{Z}) &= -\sum_{\mathbf{z}} p(\mathbf{z}) \log p(\mathbf{z}) = \\ &= -\sum_{z_1} \dots \sum_{z_K} p(z_1, \dots, z_K) \log p(z_1, \dots, z_K) \end{aligned}. \quad (8)$$

As with single-feature mutual information, the probabilities $p(\mathbf{z}_a)$, $p(\mathbf{z}_b)$ and $p(\mathbf{z}_a, \mathbf{z}_b)$ can be estimated from a multi-feature joint histogram $h(\mathbf{z}_a, \mathbf{z}_b)$. Unfortunately, even in case of two features, the four-dimensional joint histogram $h(\mathbf{z}_a, \mathbf{z}_b)$ would be so sparse that a correct estimation of joint probabilities would become practically impossible. Fortunately, by making certain assumptions on feature probability distributions, the estimation of joint probability in a high dimensional feature space can be solved efficiently.

Let the features in image feature vector $\mathbf{z}(x)$ be divided into feature $i(x)$, which we call the basic feature, and the vector $\mathbf{v}(x)$ of remaining features, which we call additional features, *i.e.*, $\mathbf{z}(x) = (i(x), \mathbf{v}(x))$ and $\mathbf{v}(x) = (v_1(x), \dots, v_{K-1}(x))$. Using the known property of entropy (Cover and Thomas 1991) that

$$H(\mathbf{Z}) = H(Z_k) + H((Z_1, \dots, Z_{k-1}, Z_{k+1}, \dots, Z_K) | Z_k), \quad (9)$$

the multi-feature mutual information defined by (Eq. 6) may be decomposed into mutual information $MI(I_a, I_b)$ of basic features $i_a(x)$ and $i_b(x)$, and additional information $AI(I_a, I_b, \mathbf{V}_a, \mathbf{V}_b)$ that is contributed by additional features $\mathbf{v}_a(x)$ and $\mathbf{v}_b(x)$

where $H(\mathbf{V}_a|I_a)$, $H(\mathbf{V}_b|I_b)$ and $H(\mathbf{V}_a, \mathbf{V}_b|I_a, I_b)$ are conditional entropies of \mathbf{V}_a , \mathbf{V}_b and $(\mathbf{V}_a, \mathbf{V}_b)$, respectively.

If basic features $i_a(x)$ and $i_b(x)$ are observed values of discrete random variables I_a and I_b then the conditional entropies $H(\mathbf{V}_a|I_a)$, $H(\mathbf{V}_b|I_b)$ and $H(\mathbf{V}_a, \mathbf{V}_b|I_a, I_b)$ can be obtained as (Cover and Thomas 1991)

$$H(\mathbf{V} | I) = \sum_i p(i)H(\mathbf{V} | i), \quad (12)$$

$$H(\mathbf{V} | i) = -\int p(\mathbf{v} | i) \log p(\mathbf{v} | i) d\mathbf{v}, \quad (13)$$

where $p(i)$ stands for probabilities $p(i_a)$, $p(i_b)$ or $p(i_a, i_b)$, and $H(\mathbf{V}|i)$ for entropies $H(\mathbf{V}_a|i_a)$, $H(\mathbf{V}_b|i_b)$ or $H(\mathbf{V}_a, \mathbf{V}_b|i_a, i_b)$ of conditional distributions $\mathbf{V}_a|i_a$, $\mathbf{V}_b|i_b$ or $\mathbf{V}_a, \mathbf{V}_b|i_a, i_b$, respectively, while $p(\mathbf{v}|i)$ stands for conditional probability distributions $p(\mathbf{v}_a|i_a)$, $p(\mathbf{v}_b|i_b)$ or $p(\mathbf{v}_a, \mathbf{v}_b|i_a, i_b)$. The additional information $AI(I_a, I_b, \mathbf{V}_a, \mathbf{V}_b)$ is thus defined by conditional joint probability distributions $p(\mathbf{v}_a|i_a)$, $p(\mathbf{v}_b|i_b)$ and $p(\mathbf{v}_a, \mathbf{v}_b|i_a, i_b)$.

Suppose that the conditional distributions $\mathbf{V}_a|i_a$, $\mathbf{V}_b|i_b$ and $\mathbf{V}_a, \mathbf{V}_b|i_a, i_b$, are approximately normal. If random variables in feature vector \mathbf{V} are distributed normally, then their joint probability $p(\mathbf{v})$ may be defined by the vector of mean values $\boldsymbol{\mu}_v$ and the covariance matrix $\boldsymbol{\Sigma}_v$. Furthermore, the entropy of \mathbf{V} can be easily assessed (Cover and Thomas, 1991) from the determinant of the covariance matrix $\boldsymbol{\Sigma}_v$ as

$$H(\mathbf{V}) = \frac{1}{2} \log |\boldsymbol{\Sigma}_v| + \frac{n}{2} \log(2\pi e). \quad (14)$$

where n is the dimension of $\boldsymbol{\Sigma}_v$. The conditional entropies $H(\mathbf{V}_a|i_a)$, $H(\mathbf{V}_b|i_b)$ and $H(\mathbf{V}_a, \mathbf{V}_b|i_a, i_b)$ may be thus defined by covariance matrices $\boldsymbol{\Sigma}_{v_a|i_a}$, $\boldsymbol{\Sigma}_{v_b|i_b}$ and $\boldsymbol{\Sigma}_{v_a, v_b|i_a, i_b}$, of conditional distributions $\mathbf{V}_a|i_a$, $\mathbf{V}_b|i_b$ and $\mathbf{V}_a, \mathbf{V}_b|i_a, i_b$, respectively. In conclusion, the estimation of conditional entropies $H(\mathbf{V}|i)$ for each intensity value i (Eq. 13) through covariance matrices requires much fewer samples than the estimation through high dimensional histograms, which makes the proposed approach more effective. Conditional entropies $H(\mathbf{V}|i)$ of all intensity values i are used to estimate conditional entropies $H(\mathbf{V}|I)$ (Eq. 12), defining the information $AI(I_a, I_b, \mathbf{V}_a, \mathbf{V}_b)$.

EXPERIMENTAL DATA SETS

Two different image data sets were used to evaluate the proposed multi-feature mutual information similarity measure. The first data set consisted of

MR-T1 and CT images of vertebra L3 of a lumbar spine phantom (Fig. 1), which was originally designed to evaluate 3D/2D registrations (Tomažević *et al.*, 2003). The MR and CT images were acquired with intra-slice resolutions of 0.39 x 0.39 mm and 0.27 x 0.27 mm and inter-slice resolutions of 1.9 mm and 1 mm, respectively. A sub-volume, containing the whole vertebra L3 and some of the two neighboring vertebrae was manually defined for each MR and CT image. The sizes of MR and CT sub-volumes were 238 x 238 x 22 and 341 x 340 x 43 voxels, respectively. The “gold standard” rigid registration was established by minimizing the distance between the centers of six fiducial markers rigidly attached to the spine phantom and visible on both MR and CT images. The accuracy of “gold standard” registration was obtained by estimating the target registration error (TRE) from the residual of fiducial marker registration (Fitzpatrick *et al.*, 1998). For vertebra L3, the expected TRE of “gold standard” MR to CT image registration was estimated to be 0.31 mm. A detailed description of image acquisition and “gold standard” registration can be found in Tomažević *et al.* (2004). To evaluate the performance of the proposed registration algorithm in case of poor image quality, the MR image was artificially corrupted by adding intensity inhomogeneity. Intensity inhomogeneity was simulated by changing the original intensities $i(x)$ to $i'(x)$ by applying the following inhomogeneity model

$$i'(x) = i(x)e^{-\varepsilon|x-x_0|^2} \quad (15)$$

where x_0 was the spatial position where inhomogeneity was centered and ε was the inhomogeneity scale factor. The position x_0 was set at the image corner $x_0 = (0,0,0)$, while four different scale factors ε were implemented ($\varepsilon = 5 \cdot 10^{-5}$, $15 \cdot 10^{-5}$, $25 \cdot 10^{-5}$, $35 \cdot 10^{-5}$) to obtain MR images with different levels of inhomogeneity (Fig. 2).

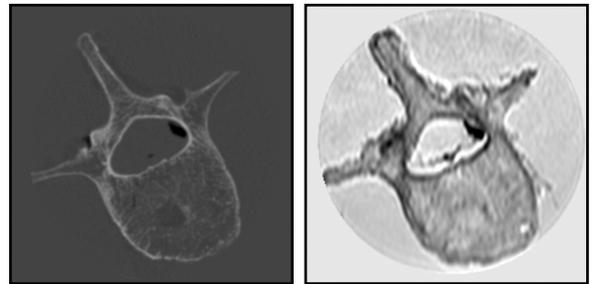


Fig. 1. A cross-section of CT (left) and MR-T1 (right) image of lumbar vertebra L3.

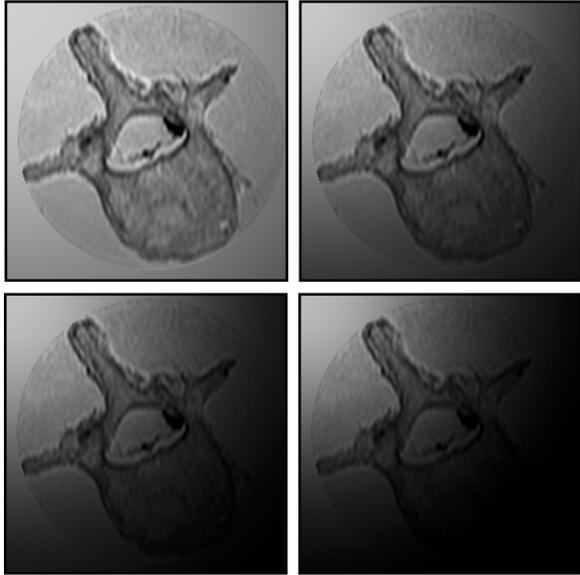


Fig. 2. Cross-sections of the MR-T1 image of lumbar vertebra L3 artificially corrupted by different levels of intensity inhomogeneity $\varepsilon = 5 \cdot 10^{-5}$ (top left), $\varepsilon = 15 \cdot 10^{-5}$ (top right), $\varepsilon = 25 \cdot 10^{-5}$ (bottom left), and $\varepsilon = 35 \cdot 10^{-5}$ (bottom right).

The second image data set consisted of PET images of a human head and corresponding proton density (PD), T1 relaxation time (T1), and T2 relaxation time (T2) magnetic resonance (MR) images (Fig. 3).

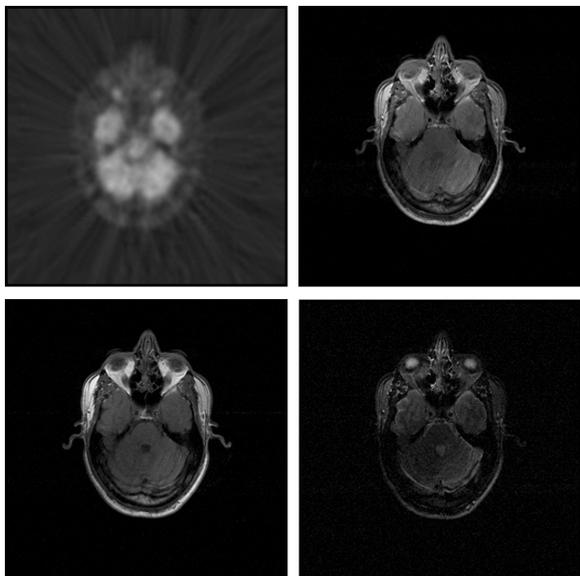


Fig. 3. Cross-sections of corresponding PET (top left), PD (top right), T1 (bottom left) and T2 (bottom right) images from the RIRE project.

The images of heads of seven different patients were acquired as part of Retrospective Image Registration Evaluation (RIRE) study (West *et al.*, 1997). Six subsets of MR images were available, with

four to seven images in each subset. Three subsets consisted of raw images of each MR modality (PD, T1, T2) and three subsets (PDr, T1r, T2r) consisted of corresponding images, corrected (rectified) for scanner-dependent geometric distortions. The size of PET images was $128 \times 128 \times 15$ voxels. The inter-slice distance of PET images was 8 mm, and the intra-slice resolution was 2.59×2.59 mm, except of one image, which had intra-slice resolution 1.94×1.94 mm. The sizes of MR images were $256 \times 256 \times 26$ or $256 \times 256 \times 24$ voxels. The intra-slice resolution of all MR images was 1.25×1.25 mm with 4 mm of inter-slice distance. Gold standard registrations, not revealed to participants of RIRE, between PET and MR images were obtained by means of fiducial markers implanted into the patient's skull.

FEATURE EXTRACTION AND SIMILARITY MEASURE OPTIMIZATION

For calculating the proposed multi-feature mutual information, image intensities of floating and reference images were used as basic features, $i_a(x)$ and $i_b(x)$, while gradients of image intensities were used as additional features, $\mathbf{v}_a(x)$ and $\mathbf{v}_b(x)$, each containing gradient components in x , y , and z directions. The original images used in our experiments had highly anisotropic image voxel sizes. The frequency of intensity signal in the intra-slice direction was much higher than the frequency in inter-slice direction. To correct for the anisotropic voxel sizes and different frequencies, each image was first isotropically resampled by linear interpolation. For MR-CT registrations the MR and CT images were resampled to 0.954 mm and 0.526 mm voxel sizes, respectively, while for PET-MR registrations images were resampled to voxel sizes equal to the smallest voxel dimension before resampling. The basic features $i_a(x)$ and $i_b(x)$ used for calculating the single-feature and multifeature mutual information measures were obtained directly from intensities of resampled images. To reduce image noise and obtain similar magnitudes of intra-slice and inter-slice intensity gradients, the gradients were obtained by implementing a simple symmetric gradient operator after convolving intensities with a Gaussian kernel of scale σ

$$\mathbf{v}(x) = \nabla(i(x) * g(x, \sigma)). \quad (16)$$

The choice of scale σ should assure that inter-slice gradients are suppressed while intra-slice gradients are not affected by filtering. The sizes of convolution kernels were chosen to be $1/3$ - $1/4$ of the average intra-slice resolution of two images involved in registration, and were 0.5 mm and 1.5 mm for MR-CT and PET-

MR experiments, respectively. Image intensities were discretized into M , $M=256$, 64 and 16 intensity levels. In both registration experiments, Powell's optimization method (Press *et al.*, 1992) was used to optimize the chosen similarity measure for six rigid-body transformation parameters $(t_x, t_y, t_z, \omega_x, \omega_y, \omega_z)$.

RESULTS

MR-CT REGISTRATION

To test the accuracy and reliability of the multi-feature mutual information similarity measure and compare it to the single-feature mutual information, registrations were performed from twenty different starting positions around the "gold standard" registration for each MR-CT data set and each similarity measure. The starting positions of the floating MR image with respect to the registered reference CT image were randomly generated by alternating three translation parameters (from 0-8mm) and three rotation parameters (from 0-22.9°) in such a way that the distance from "gold standard" position in normalized space of rigid transformations was uniformly distributed on the interval of 0-8mm (0-22.9°). The following assumption was used to normalize the parametrical space of rigid transformation: rotation of a volume containing a single-vertebra of size 80 mm around its center for 0.1 radians (5.7°) causes mean translation of vertebra points of 2 mm (Tomažević *et al.* 2003). For each individual registration, the registration error was calculated as the root mean square (rms) target registration error (TRE). The positions of CT image voxels belonging to bone and obtained by bone thresholding were chosen as target points. A registration was treated as successful if the obtained rmsTRE was below 1.9 mm, which was the largest voxel size of images involved in registrations. The mean of rmsTRE of all twenty starting positions, none of which was smaller than 1.9 mm, was 4.3 mm with a standard deviation of 1.62 mm. The results for MR-

CT registrations, using the two similarity measures, five different intensity inhomogeneities, and $M=256$, 64 and 16 intensity levels are presented in Table 1. Table 1 also gives the percentage of misregistration and the mean and standard deviation of registration errors of all successful registrations. In addition, Table 1 gives the significance ($P=0.01$) of a two tailed hypothesis test on differences of mean registration errors of two similarity measures. Results in Table 1 indicate, that by increasing the intensity inhomogeneity the registration error and percentage of misregistrations increase much more when the single-feature mutual information is used. This is true regardless of the number of intensity levels.

RIRE PET-MR REGISTRATION

In these experiment 35 combinations of PET-MR images were rigidly registered by the standard single-feature mutual information of image intensities and the proposed multi-feature mutual information of image intensities and image intensity gradients as similarity measures. For both similarity measures, the starting position for each individual PET-MR registration was obtained by aligning the centers of the floating PET image and reference MR image, while the number of intensity levels was set to 64. The obtained transformations were submitted to the authors of RIRE project, and in return, they provided the registration errors for a particular registration. A PET-MR registration was considered successful if the average TRE error over the volumes of interest was smaller than 8 mm, which was the largest voxel dimension of the MR or PET image. A summary of registration errors for both mutual information measures and for each PET-MR subset appears in Table 2. All median and maximum registration errors, except for PET-T1r were smaller for the multi-feature mutual information than for the single-feature mutual information. Multi-feature mutual information also outperformed the classical mutual information when the number of misregistrations was taken into account.

Table 1. Results of MR to CT image registrations with the single-feature mutual information (SF) and multi-feature mutual information (MF) similarity measure for different number of intensity levels and different levels ε of intensity inhomogeneity in the MR image. (-) indicates that none of the registrations was successful.

	Registration errors (mm) [Mean (Std)]		Misregistrations (%)		Significance (P=0.01) *	
	SF	MF	SF	MF	$\mu_{SF} < \mu_{MF}$	$\mu_{SF} > \mu_{MF}$
256 intensity levels						
$\varepsilon=0$	0.84 (0.027)	0.90 (0.045)	0	15	*	
$\varepsilon=5 \cdot 10^{-5}$	1.06 (0.018)	0.85 (0.117)	0	5		*
$\varepsilon=15 \cdot 10^{-5}$	1.34 (0.028)	0.97 (0.070)	50	25		*
$\varepsilon=25 \cdot 10^{-5}$	-	0.90 (0.038)	100	25		*
$\varepsilon=35 \cdot 10^{-5}$	-	0.99 (0.045)	100	40		*
64 intensity levels						
$\varepsilon=0$	0.98 (0.018)	0.84 (0.010)	0	0		*
$\varepsilon=5 \cdot 10^{-5}$	1.11 (0.018)	0.89 (0.010)	0	0		*
$\varepsilon=15 \cdot 10^{-5}$	1.22 (0.042)	0.92 (0.008)	25	35		*
$\varepsilon=25 \cdot 10^{-5}$	1.45 (-)	0.90 (0.020)	95	40		*
$\varepsilon=35 \cdot 10^{-5}$	-	0.97 (0.012)	100	25		*
16 intensity levels						
$\varepsilon=0$	1.04 (0.015)	0.79 (0.020)	0	0		*
$\varepsilon=5 \cdot 10^{-5}$	1.11 (0.016)	0.85 (0.019)	0	0		*
$\varepsilon=15 \cdot 10^{-5}$	1.38 (0.132)	0.98 (0.010)	40	15		*
$\varepsilon=25 \cdot 10^{-5}$	1.89 (-)	0.99 (0.016)	95	35		*
$\varepsilon=35 \cdot 10^{-5}$	-	0.92 (0.023)	100	40		*

Table 2. Registration results for single-feature (SF) and multi-feature (MF) mutual information registration of PET and MR images.

	Registration errors (mm)				Misregistrations		Registration experiments
	Median		Maximum		SF	MF	
	SF	MF	SF	MF			
PET-T1	4.17	4.13	196.98	8.82	2	-	7
PET-T2	4.21	2.23	80.91	4.54	2	-	7
PET-PD	4.47	3.39	52.10	12.66	2	-	7
PET-T1r	2.25	3.24	4.78	7.60	-	-	4
PET-T2r	2.23	2.17	5.40	4.72	-	-	5
PET-PDr	3.34	3.04	130.28	4.79	2	-	5

COMPARISON WITH OTHER REGISTRATION METHODS IN THE RIRE PROJECT

Fig. 4 summarizes the results of various PET to MR registration methods tested in the RIRE project and compares these results with the results of the proposed multi-feature mutual information similarity measure based on image intensities and intensity gradients. Fig. 4 comprises registration results of those registration methods published on the RIRE web page which contain complete results, *i.e.*, results of all patients in a data subset. Fig. 4 shows results for six PET-MR data subsets. Each dot in a graph

represents the median and maximum registration errors for a particular method. The results of the proposed multi-feature mutual registration method are illustrated by dotted lines. Three different implementations of mutual information based registration, *i.e.*, the methods of Collignon *et al.* (1995), Studholme *et al.* (1996), and Thevenaz and Unser (2000), are especially marked as CO, HI and TH, respectively. In Fig. 4, the absence of TH result for PET-T1 registration is due to TH maximum error, which exceeds the graph's boundaries, while the results of CO are excluded from PET-T1r results because the authors did not report the registration results for the complete PET-T1r data set.

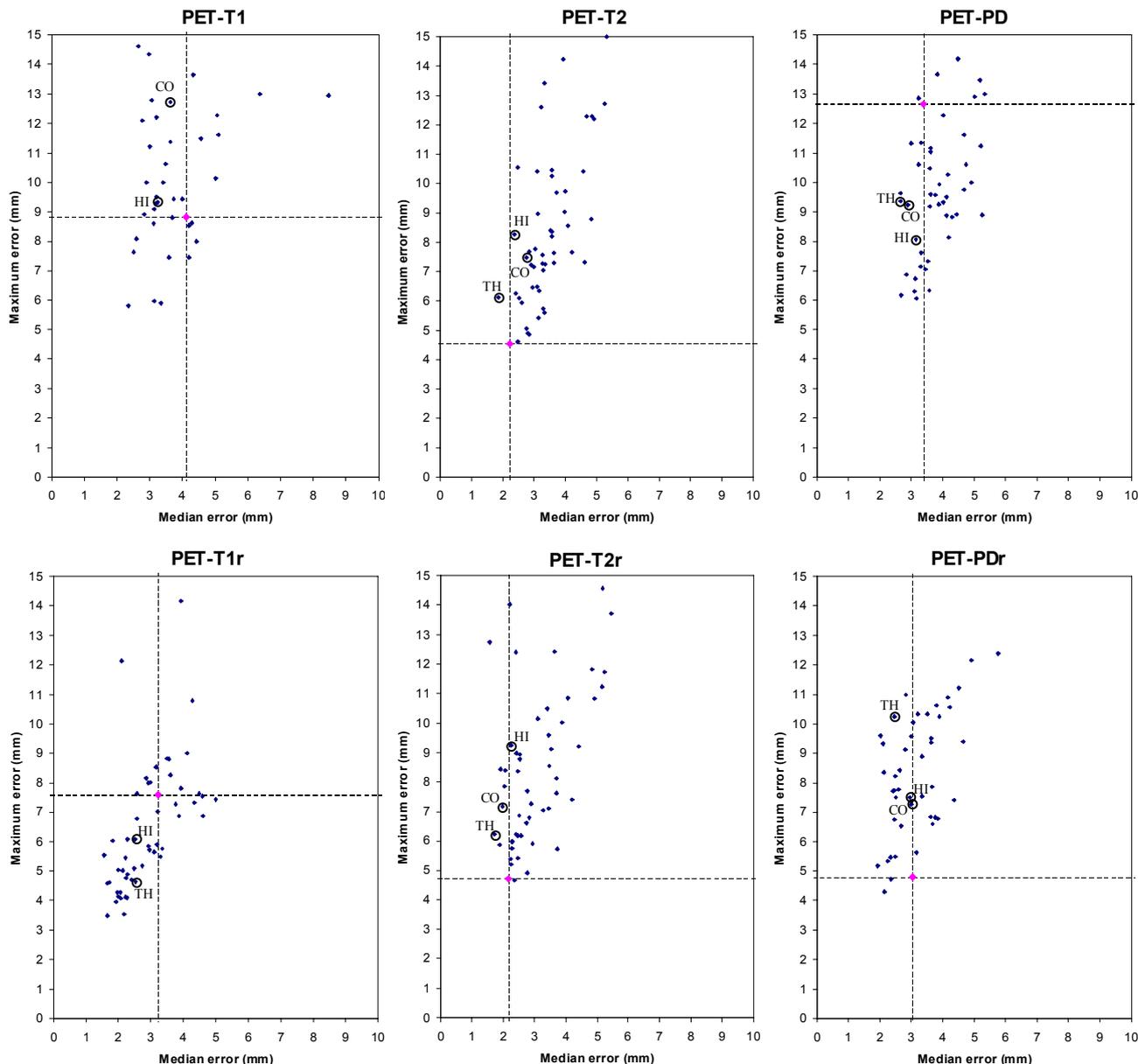


Fig. 4. Performance of the proposed multi-feature mutual information (dotted lines) and other registration methods (dots) in the RIRE study. For a PET-MR data subset, each dot represents the median and maximum errors of a particular registration method.

DISCUSSION

In the last decade, information-theoretic similarity measures, especially mutual information and its derivatives, have been extensively used for rigid and non-rigid, mono- and multi-modal image registrations. However, in cases of low global intensity feature correspondence these measures are not robust enough. The robustness can be improved by adding spatial information in the form of local intensity feature changes to the global intensity correspondence. In this paper we have showed how spatial information in

the form of intensity gradients may be included in information-theoretic similarity measures. The proposed measure was evaluated by registering intensity inhomogeneity corrupted MR images with CT images. In addition, the measure was tested for the registration of PET to MR images from the widely used Retrospective Image Registration Evaluation project image database.

The aim of the MR to CT registration experiments was to establish and compare the performances of standard single-feature mutual information and the proposed multi-feature mutual information similarity

measure in the presence of intensity inhomogeneities, which is an important problem in MR imaging. Intensity inhomogeneity is primarily caused by variation in the deposition of energy into tissue and the reception of energy from tissue by the radiofrequency coil. The problem gets worse with increasing magnet strength and it is especially bad in the new 7T MR scanners. Intensity inhomogeneity is visibly noticeable as gradual spatial intensity changes that are difficult to correct by retrospective shading correction methods (Vovk *et al.*, 2007). Results in Table 1 show, that by increasing the intensity inhomogeneity of an MR image, the single-feature mutual information registration error increases significantly, regardless of the number of intensity levels, while the multi-feature mutual information registration error is significantly lower and much less affected by intensity inhomogeneities. When intensity inhomogeneity is increased, the percentage of misregistrations increases for both similarity measures, but the use of single-feature mutual information results in a higher percentage of misregistrations and fails completely at intensity inhomogeneity $\varepsilon = 25 \cdot 10^{-5}$. On the contrary, the multi-feature mutual information is successful in at least 60% of registrations at intensity inhomogeneities $\varepsilon = 25 \cdot 10^{-5}$ and $35 \cdot 10^{-5}$. Finally, multi-feature mutual information results in significantly smaller registration errors than single-feature mutual information, except when registering an inhomogeneity free image and using 256 intensity levels.

The summary of the RIRE PET-MR registration experiment (Table 2) shows that median TREs were smaller for multi-feature mutual information criterion than for standard mutual information, except for PET-T1r registrations where the median and maximum TREs were somewhat larger. Furthermore, large maximum registration errors obtained by the standard mutual information suggest that some of the obtained registered positions did not come even close to true registration, which was reflected in the number of misregistrations. Moreover, for misregistrations we observed that the size of image overlap was significantly smaller after registration than before. This suggests that misregistrations were probably caused by a possible effect of the size of image overlap on mutual information, *i.e.*, smaller image overlap probably increases the mutual information despite the fact that the image transformation is moving away from the true registration solution. The effect of the size of image overlap was stronger when there was little information between images, which is true for PET to MR registrations.

When comparing the registration results of the proposed multi-feature mutual information similarity measure with other methods in the RIRE project (Fig. 4), the median registration error may be an indication of the accuracy of a registration method, while the maximum error points to the robustness of a method. When taking into account the median and maximum registration errors, it is clear that the results of the proposed method outperformed the results of other methods for PET-T2 and PET-T2r registrations. For PET-T2 registration, our method achieved the lowest maximum registration error and second lowest median error, while for PET-T2r seven methods showed better median errors but were less robust in comparison to our method. Similarly, for PET-PDr registration our method provided one of the lowest maximum registration errors. Unfortunately, this cannot be claimed for PET-T1, PET-T1r and PET-PD registrations, where our method was less successful compared to other methods. However, it should be emphasized that some methods were especially designed for PET to MR registrations of the human head. Therefore, we further focused our comparisons on mutual information based registration methods. While the method of Studholme *et al.* (1996) is similar to the one proposed by Collignon *et al.* (1995), the method of Thevenaz and Unser (2000) implements an improved multi-resolution approach, based on B-spline image intensity interpolation and optimization, employing analytically obtained gradients of the similarity measure. The median errors of CO, HI, TH methods and our method were within 1 mm. The median error produced by the TH method was slightly lower than ours, most probably due to the more sophisticated interpolation scheme used by TH. In addition, our method resulted in smaller maximum registration errors for four out of six PET-MR data sets. Its superiority was especially obvious in PET-T2, PET-T2r and PET-PDr registrations.

In this work we have proposed an extension of the standard single-feature mutual information similarity measure to a multi-feature mutual information measure. The proposed multi-feature mutual information strictly follows the information theory and solves the problem of efficient estimation of multi-feature mutual information from sparse high-dimensional histograms. Estimation is comprised of the classical approach to mutual information calculation, which is conducted on joint histograms of basic features and the estimation of additional information from conditional covariance matrices of additional features. Such an approach is more efficient and applicable to an arbitrary

number of additional features as long as they are approximately normally distributed. As presented, multi-feature mutual information registration approach is general. The approach is feasible when additional features will provide additional information on correspondence between two images that are registered. If features from one image have large correspondence with features on the other image, then their contribution to multi-feature mutual information will be larger than for features with lower correspondence. If features between images have no correspondence, their contribution to similarity measure will be insignificant. Furthermore, the contribution of features that are highly correspondent with other features within a single image will be irrelevant since those features are redundant. Rigid registration experiments indicate that the multi-feature mutual information outperforms the single-feature mutual information measure. The contribution of additional image features is especially significant in cases of poor intensity dependences.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Higher Education, Science and Technology, of the Republic of Slovenia under the grants P2-0232, L2-7381, L2-9758 and Z2-9366. The images and the standard transformation(s) were provided as part of the project, "Retrospective Image Registration Evaluation", National Institutes of Health, Project Number 8R01EB002124-03, Principal Investigator, J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN, USA.

REFERENCES

- Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G (1995). Automated multi-modality image registration based on information theory. *Information Processing in Medical Imaging*. Norwell, MA: Kluwer, 263-74.
- Cover TM, Thomas JA (1991). *Elements of information theory*. New York, USA: John Wiley & Sons.
- Fitzpatrick JM, West JB, Maurer CR, Jr. (1998). Predicting error in rigid-body point-based registration. *IEEE Trans Med Imag* 17:694-702.
- Hawkes DJ (1998). Algorithms for radiological image registration and their clinical application. *J Anat* 193: 347-61.
- Hero AO, Ma B, Michel OJJ, Gorman J (2002). Applications of entropic spanning graphs. *IEEE Sig Proc Mag* 19: 85-95.
- Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P (1997). Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imag* 16: 187-98.
- Maes F, Vandermeulen D, Suetens P (2003). Medical image registration using mutual information. *Proceedings of the IEEE* 91:1699-722.
- Maintz JB, Viergever MA (1998). A survey of medical image registration. *Med Image Anal* 2:1-36.
- Neemuchwala H, Hero A, Carson P (2005). Image matching using alpha-entropy measures and entropic graphs. *Sig Proc* 85:277-96.
- Neemuchwala H, Hero A, Zabuawala S, Carson P (2006). Image registration methods in high-dimensional space. *Int J Imaging Syst Technol* 16:130-45.
- Pluim JP, Maintz JB, Viergever MA (2000a). Image registration by maximization of combined mutual information and gradient information. *IEEE Trans Med Imag* 19: 809-14.
- Pluim JP, Maintz JB, Viergever MA (2000b). Interpolation artefacts in mutual information-based image registration. *Computer Vision and Image Understanding* 77:211-32.
- Pluim JP, Maintz JB, Viergever MA (2003). Mutual-information-based registration of medical images: a survey. *IEEE Trans Med Imag* 22:986-1004.
- Press WH, Flannery BP, Teukolosky SA, Vetterling WT (1992). *Numerical recipes in C*. Cambridge, UK: Cambridge University Press.
- Roche A, Malandain G, Ayache N (2000). Unifying maximum likelihood approaches in medical image registration. *Int J Imaging Syst Technol* 11:71-80.
- Rodríguez-Carranza CE, Loew MH (2000). A weighted and deterministic entropy measure for image registration using mutual information. *Medical Imaging:Image Processing*. Hanson KM. Bellingham, WA, SPIE. 3338: 155-66.
- Rueckert D, Clarkson MJ, Hill DLG, Hawkes DJ (2000). Non-rigid registration using high-order mutual information. *Medical Imaging:Image Processing*. Hanson KM. Bellingham, WA, SPIE. 3979:438-47.
- Sabuncu MR, Ramadge P (2008). Using spanning graphs for efficient image registration. *Image Processing, IEEE Transactions on* 17:788-797.
- Sabuncu MR, Ramadge PJ (2003). Spatial information in entropy-based image registration. *Biomedical Image Registration* 2717:132-41.
- Staring M, van der Heide UA, Klein S, Viergever MA, Pluim J (2009). Registration of cervical MRI using multifeature mutual information. *Medical Imaging, IEEE Transactions on* 28:1412-21.
- Studholme C, Hill DL, Hawkes DJ (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition* 32:71-86.
- Studholme C, Hill DLG, Hawkes DJ (1996). Automated 3D registration of MR and CT images of the head. *Med Image Anal* 1:163-75.
- Špiclin Ž, Likar B, Pernuš F (2012). Groupwise registration of multi-modal images by an efficient joint entropy minimization scheme. *Image Processing, IEEE Transactions on* In Press.

- Thevenaz P, Unser M (2000). Optimization of mutual information for multiresolution image registration. *IEEE Trans Image Proc* 9:2083-99.
- Tomažević D, Likar B, Pernuš F (2004). "Gold standard" data for evaluation and comparison of 3D/2D registration methods. *Comput Aided Surg* 9:137-44.
- Tomažević D, Likar B, Slivnik T, Pernuš F (2003). 3-D/2-D registration of CT and MR to X-ray images. *IEEE Trans Med Imag* 22:1407-16.
- Tsao J (2003). Interpolation artifacts in multimodality image registration based on maximization of mutual information. *IEEE Trans Med Imag* 22(7):854-64.
- Viola P, Wells WM III, (1995). Alignment by maximization of mutual information. *Proc. 5th Int. Conf. Computer Vision*. Boston, MA:16-23.
- Vovk U, Pernuš F, Likar B (2007). A review of methods for correction of intensity inhomogeneity in MRI. *Medical Imaging, IEEE Transactions on* 26:405-21.
- Wells WM III, Viola P, Atsumi H, Nakajima S, Kikinis R (1996). Multi-modal volume registration by maximization of mutual information. *Med Image Anal* 1:35-51.
- West J, Fitzpatrick JM, Wang MY, Dawant BM, Maurer CR, Jr., Kessler RM, *et al.* (1997). Comparison and evaluation of retrospective intermodality brain image registration techniques. *J Comput Assist Tomogr* 21: 554-66.