Point Similarity Measures for Non-Rigid Registration of Multi-Modal Data

Peter Rogelj, Stanislav Kovačič
University of Ljubljana, Faculty of Electrical Engineering,
Tržaška c. 25, 1000 Ljubljana, Slovenia
E-mail: peter.rogelj@fe.uni-lj.si, stanek@fe.uni-lj.si

and

James C. Gee
University of Pennsylvania, Department of Radiology
1 Silverstein, 3400 Spruce Street, Philadelphia, PA 19104, USA
E-mail: gee@rad.upenn.edu

Version: May 6, 2003

Running head:

Point Similarity Measures

Correspondence:

Peter Rogelj, Faculty of Electrical Engineering, Trzaska 25, 1000 Ljubljana,

phone: +386 1 4768 876, fax: +386 1 4768 279, e-mail: peter.rogelj@fe.uni-lj.si

Abstract

High-dimensional non-rigid registration of multi-modal data requires similarity measures with two important properties: multi-modality and locality. Unfortunately all commonly used multi-modal similarity measures are inherently global and cannot operate on small image regions. In this paper we propose a new class of multi-modal similarity measures, which are constructed from information of the whole images but can be applied pointwise. Due to their capability of measuring correspondence for individual image points we call them point similarity measures. Point similarity measures can be derived from global measures and enable detailed relative comparison of local image correspondence. We present a set of multi-modal point similarity measures based on joint intensity distribution, and test them as an integral part of non-rigid multi-modal registration system. The comparison results show that segmentation based measure, which models the joint distribution as a sum of intensity classes, performs best. When intensity classes do not exist or cannot be accurately modeled, each intensity pair can be treated as a separate class, which results in a more general measure, suitable for various non-rigid registration tasks.

Key Words: similarity measure, point similarity, multi-modality, non-rigid registration.

1. INTRODUCTION

Multi-modal registration and non-rigid registration have mainly been considered as two different and independent problems. Evolution of non-rigid registration has focused on development of algorithms for high-dimensional warping, but remained limited to mono-modal data. On the other hand evolution of multi-modal registration has been mainly concentrated on development of multi-modal similarity measures and parametric optimization techniques, while limited to low-dimensional transformations, e.g. rigid or affine.

Recently there have also been efforts to develop algorithms with both multi-modal and non-rigid properties. Such algorithms enable combining information of different modality images of body parts that undergo non-rigid motion between image acquisitions. They can be used for various medical applications, e.g. development of anatomical atlases, comparison of subjects, medical examination, planing and inspection of medical treatment, intraoperative localization, etc. In general, registration can be landmark based or intensity based, for the review of registration procedures see [24, 37]. Landmark based approaches [32, 27] involve extraction of landmarks that need to matched. By interpolating the discrete matching of the landmarks, one tries to obtain a dense mapping for the whole image. Intensity based approaches use the information of the whole images and find the dense mapping by optimizing some criterion function that incorporates a measure of image similarity. In this paper we focus on intensity based approaches, which form an active area of research in the field of non-rigid multi-modality registration. We classify these approaches into four categories.

The first category is registration based on global similarity measures [33, 25, 31], which detect improvement of local image correspondence by observing global image similarity. These approaches are appropriate for detecting simple smooth deformations, which can be modeled by low parameter deformation models. When models with large number of parameters are used these approaches became very time consuming, as every estimation of local properties requires recomputation of similarity of the whole images.

The second category of registration algorithms comprises block matching techniques [5, 10, 14, 17, 20]. They divide a source image into smaller regions that are independently registered to target image using multiple low-dimensional transformations. Region registration results are then used to obtain a smooth global transformation. In general, registration results improve by reducing region sizes and thus increasing the number of regions. However, regions cannot be arbitrarily small, as reduction of region size decreases performance of multi-modal similarity

measures that are used for estimating image correspondence.

Approaches in the third category are based on intensity transformation. One such approach to the problem of precise non-rigid registration was proposed by Guimond et al. [13]. The method is based on applying intensity transformation to one of the images, so that it matches the intensity properties of the other image. After that, images can be registered using high-dimensional mono-modal, instead of multi-modal, non-rigid registration algorithm. Such approaches are suitable for images with functional intensity dependency.

Finally, the fourth category of registration approaches directly follows high-dimensional mono-modal solutions. This strategy requires the use of similarity measures that can estimate multi-modal correspondence of small image regions. The size of image regions is crucial, as smaller regions enable detection of more local image differences, enabling better correction and therefore better registration of highly deformed images. However, matching of small image regions, e.g. individual voxels, is ill-posed [10, 23], if displacements are calculated independently for each region. To regularize the registration, voxel interdependencies must be introduced by spatial deformation model, based on elasticity, fluid mechanics or any other physical or fictitious continuum mechanics property. Due to the use of small image regions these approaches provide good control over spatial deformation properties and enable correction of most localized image discrepancies.

To make the last category of registration approaches possible, image properties must be measured for small image regions. This requires multi-modal similarity measures that can be used locally. We propose a new class of multi-modal similarity measures, which can measure properties of arbitrary small image regions, including the smallest meaningful image regions that cover only individual voxels. The problem of locality of multi-modal similarity measures is highlighted in the following subsection. The proposed solution is described in Section 2.

1.1. Locality and multi-modality of similarity measures

A variety of similarity measures exist [15, 16], and were mostly developed for specific image registration tasks. In general they can be divided into mono-modal measures, which are appropriate for comparing and registering images of similar intensity characteristics, and multi-modal measures, which can deal with complex image intensity relations that appear in cross-modality registration. Similarity measures can also be divided into local and global measures, but this classification is not clear unless the term locality is strictly defined. Given a local similarity measure, its global equivalent can be derived by summing up over the whole image. On the

opposite, global similarity measures can be used to measure local image correspondences, but the performance of most similarity measures gradually decreases as the region size decreases. The minimum region size depends on the nature of the similarity measure in a way that more general measures, which have more free parameters, require larger regions.

The simplest similarity measures are based on difference of scalar intensity, e.g. mean absolute difference $MAD(A, B)$ or mean squared difference $MSD(A, B)$,

$$MAD(A, B) = \frac{1}{N} \sum_v |i_A(v) - i_B(v)|, \tag{1}$$

$$MSD(A, B) = \frac{1}{N} \sum_v (i_A(v) - i_B(v))^2, \tag{2}$$

where $i_A(v)$ and $i_B(v)$ are image intensities in images A and B at position of voxel v, and N is the number of overlapping voxels. These measures do not tolerate any difference in brightness or contrast, but they do allow the use of arbitrarily small image regions, including individual voxels ($N = 1$).

Measures that can deal with certain image differences require larger regions. Correlation based measures, which are the most commonly used mono-modal similarity measures, belong to that group too. Such measure is the correlation coefficient $CC(A, B)$ [3],

$$CC(A, B) = \frac{1}{N} \frac{\sum_v (i_A(v) - \overline{i_A})(i_B(v) - \overline{i_B})}{(\sum_v (i_A(v) - \overline{i_A})^2 \sum_v (i_B(v) - \overline{i_B})^2)^{\frac{1}{2}}}. \tag{3}$$

Here, $\overline{i_A}$ and $\overline{i_B}$ denote mean image intensity values. Correlation coefficient assumes linear intensity relationship and therefore it can deal with differences in image contrast and brightness. Two unknown parameters of linear intensity relationship (brightness and contrast) require at least three samples (voxels). In practice $3 \times 3 \times 3$ voxel regions or larger regions are used.

The minimum region size requirement becomes much more problematic in the case of cross-modality registration, when the images are acquired using different imaging procedures. In such multi-modal cases the relation between image intensities can be quite complex and is generally not known. Multi-modal measures must be able to deal with such complex relationships. The most widely used multi-modal similarity measures are mutual information $MI(A, B)$ [38, 8] and normalized mutual information $NMI(A, B)$ [35],

$$MI(A, B) = H(A) + H(B) - H(A, B), \tag{4}$$

$$NMI(A, B) = \frac{H(A) + H(B)}{H(A, B)}. \tag{5}$$

Here, $H(A)$, $H(B)$ and $H(A,B)$ are marginal and joint entropies,

$$H(A) = -\sum_{i_A} p(i_A) \log p(i_A), \tag{6}$$

$$H(B) = -\sum_{i_B} p(i_B) \log p(i_B), \tag{7}$$

$$H(A,B) = -\sum_{\mathbf{i}} p(\mathbf{i}) \log p(\mathbf{i}), \tag{8}$$

where $\mathbf{i}$ denotes an intensity pair $[i_A, i_B]$, $p(i_A)$, $p(i_B)$ are marginal intensity probabilities and $p(\mathbf{i}) = p(i_A, i_B)$ is a joint intensity probability, estimated from the images. Mutual information measures actually measure statistical dependence between the image intensities, to estimate how much one image tells about the other one. But when the image regions used to estimate the joint distribution are small, its statistical significance is low and the similarity measures poorly represent actual image correspondence.

Various solutions have been proposed in order to improve locality of multi-modal measures. The most obvious solution is intensity binning, decreasing the number of intensity bins [21]. This directly improves the estimation of joint distribution, but on the other hand it lowers the intensity sensitivity, e.g. different features that are represented by similar intensity values can be grouped into the same bin.

Another approach is to use Parzen window estimation of joint distribution [26]. It can be efficiently implemented as post-filtering of normalized joint histogram. This method lowers intensity sensitivity too. Another way to improve locality of multi-modal measures is to resort to one dimensional statistics, e.g. $H(A-B)$ or $H(A+B)$ [4], which also reduce the number of intensity bins.

Locality can also be improved by using prior information given in a form of joint distribution p_{prior} of correctly registered images. Such a solution was used in the log likelihood similarity measure proposed by Leventon and Grimson [19]. Similarity can also be calculated from a weighted sum of prior probability p_{prior} that is given in advance, and probability p_{image}, which is estimated from the images:

$$p = \lambda p_{image} + (1-\lambda) p_{prior}. \tag{9}$$

This enables additional decreasing of region size [21]. As correct prior probability is rarely known, Maintz et al. proposed to use a global distribution instead of prior one [23]. Their local measure is based on global conditional probability. Joint probability estimated from smaller image regions instead of the whole images is sometimes used in place of prior information [20].

In this paper we present a new class of multi-modal similarity measures, which are constructed from information of the whole images and can be applied pointwise. Due to their capability of measuring correspondence for individual image points we call them point similarity measures. In the next section we show our approach to derive multi-modal point similarity measures based on joint intensity distribution. Section 3 is devoted to high dimensional multi-modal non-rigid registration. This registration method is then used for comparison of the proposed point similarity measures in Section 4. We conclude the paper with discussion and conclusions in Section 5.

2. POINT SIMILARITY MEASURES

In this section we develop local similarity measures, which are capable of measuring image correspondences on the basis of individual voxels. Such measures are used for estimating similarity $S(v)$ for image point v according to its properties, e.g. intensity pair $\mathbf{i} = [i_A, i_B]$ present at this point, where i_A is intensity of point in target image A, and i_B is intensity of corresponding point in source image B. Due to this feature we call such measures *point similarity measures*. Likewise, function $S(\mathbf{i})$, which shows how point similarity varies as a function of intensity pair $\mathbf{i}$ is called point similarity function.

The basic requirement for a given point similarity measure S to be appropriate for non-rigid registration is that improvements of point similarities $S(v)$ contribute to the improvement of global similarity S_G between the whole images. Let $\mathbf{U}$ denote a local transformation that makes such improvement. This transformation deforms image B to image B' and therefore also changes correspondences of individual image voxels such that intensity pair $\mathbf{i}(v)$ at point v changes to $\mathbf{i}'(v)$:

$$\mathbf{U}: \ \mathbf{i}(v) \longrightarrow \mathbf{i}'(v), \ B \longrightarrow B'. \tag{10}$$

These local changes must be detected by point similarity measure, such that

$$S\left(\mathbf{i}'\left(v\right)\right) = S'\left(v\right). \tag{11}$$

Assuming that larger similarity means better correspondence, the increase of point similarities $S(v)$ must increase some global similarity $S_G(A, B)$:

$$\sum_n S'(v_n) > \sum_n S(v_n) \Longrightarrow S_G(A, B') > S_G(A, B), \tag{12}$$

where the summations run over all image voxels. According to this requirement, every local similarity measure corresponds to some global similarity measure.

If the transformation $\mathbf{U}$ changes the image correspondence at only one image voxel v_m and point similarities of all other image voxels remain unchanged, $S(v_n) = S'(v_n)$, $n \neq m$, then eq.(12) can be rewritten:

$$\sum_{n;n\neq m} S(v_n) + S'(v_m) > \sum_{n;n\neq m} S(v_n) + S(v_m) \Longrightarrow S_G(A, B') > S_G(A, B), \tag{13}$$

and therefore

$$S'(v_m) > S(v_m) \Longrightarrow S_G(A, B') > S_G(A, B). \tag{14}$$

Hence, the improvement of image correspondence at a single image point v_m improves the global image correspondence as well. This relation can be used for deriving point similarity measures from global similarity measures. Let us imagine we can change image correspondence at a single point v_m, e.g change image intensities $\mathbf{i}$ of that point. Then the change of point similarity $S(v_m)$ must correspond to the change of global similarity $S_G(A, B)$. One way to define point similarity measure is therefore the following:

$$S'(v_m) = S\left(\mathbf{i}'\left(v_m\right)\right) = C_1 \cdot S_G(A, B') + C_0; \tag{15}$$

where constants C_1 and C_0 can be chosen arbitrarily, whereas the sign of C_1 should be such that larger point similarity means better correspondence[1]. Two point measures that differ only in constants C_1 and C_0, e.g. $S'_1(v_m) = 2 \cdot S_G(A, B')$ and $S'_2(v_m) = S_G(A, B') - S_G(A, B)$, are equivalent and they both reach the optimum at the same local transformation. Thus, point similarity function $S(\mathbf{i})$ can always be estimated from global similarity S_G by changing image correspondence at single image point.

There is another way to derive point similarity measures. When global similarity measure can be calculated as a sum of local contributions $S(v)$, defined for each individual image voxel v:

$$S_G(A, B) = \sum_v S(v), \tag{16}$$

then $S(v)$ can be treated as point similarity at that voxel v. For example, mono-modal measure $MAD(A, B)$ is defined as a sum over all image voxels. The corresponding similarity function $S(\mathbf{i}) = -|i_A - i_B|$ is shown in Fig. 1. The similarity of a point pair can be always obtained from corresponding intensity pair $\mathbf{i} = [i_A, i_B]$, directly from given similarity function $S(\mathbf{i})$. Similarity $S(v)$ for point v can be estimated from corresponding intensities $\mathbf{i}(v) = [i_A(v), i_B(v)]$, such that $S(v) = S(\mathbf{i}(v))$. Similarity function $S(\mathbf{i}) = -|i_A - i_B|$ does not depend on the actual images and

[1] C_1 is negative if lower value of S_G means better correspondence.

on the level of their correspondence. This holds only for measures that presume certain image intensity dependence, e.g. $i_A = i_B$. In other cases point similarity measure and its similarity function adapt to actual image properties. For example, they may depend on actual image intensity distributions, as shown in section 2.1.

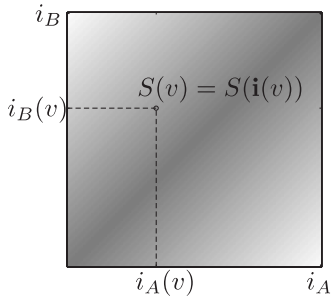


FIG. 1 Example of a point similarity function that corresponds to a MAD similarity measure. Darker color means higher similarity. Similarity of a point v can always be obtained from corresponding intensity pair $\mathbf{i}(v) = [i_A(v), i_B(v)]$.

2.1. Entropy based point similarity measure

In this section we derive point similarity measures from global entropy based similarity measures.

Let us first derive a point similarity measure from negative joint entropy $-H(A, B)$, which is the most informative part of mutual information and can be independently used as a global multi-modal similarity measure [9]. Negative sign is used such that higher similarity means better image correspondence.

Let us rewrite (8) in the following form,

$$-H(A,B) = \sum_{\mathbf{i}} p(\mathbf{i}) \log p(\mathbf{i}) = \sum_{\mathbf{i}} \frac{N_{\mathbf{i}}}{N} \log p(\mathbf{i}) = \frac{1}{N} \sum_{v} \log p(\mathbf{i}(v)), \tag{17}$$

where $N_{\mathbf{i}}$ is the number of occurrences of intensity pair $\mathbf{i}$, N is the total number of intensity pairs in the image, which equals the number of overlapping image voxels, and $\mathbf{i}(v)$ is the intensity pair located at voxel v. Note that the final summation is taken over the spatial image coordinates instead of intensities. Thus, global similarity $-H(A, B)$ can be treated as an average of local properties, defined for each voxel v.

$$-H(A,B) = \frac{1}{N} \sum_{v} S_H(v), \tag{18}$$

$$S_H(v) = \log p(\mathbf{i}(v)), \tag{19}$$

and therefore, $S_H(v)$ is a local contribution of this voxel to global joint entropy.

When transformation $\mathbf{U}$ changes image B to B', local contribution of voxel v to global joint entropy changes as well:

$$S'_H(v) = \log p'(\mathbf{i}'(v)). \tag{20}$$

Thus, voxel contributions do not depend only on transformation of that voxel, which changes intensity $\mathbf{i}(v)$ to $\mathbf{i}'(v)$, but also on transformation of the whole image, which changes joint distribution p to p'.

When performing high-dimensional non-rigid registration iteratively, in each iteration global transformation is calculated from estimated changes of all local similarities. Therefore, global transformation $\mathbf{U}$ and new global joint distribution p' are not known until similarities are estimated. So, let us express $p'(\mathbf{i}'(v))$ using Taylor expansion:

$$S'_H(v) = \log \left(p\left(\mathbf{i}'(v)\right) + \frac{\partial p(\mathbf{i}'(v))}{\partial \mathbf{U}}\mathbf{U} + \frac{1}{2}\frac{\partial^2 p(\mathbf{i}'(v))}{\partial \mathbf{U}^2}\mathbf{U}^2 + ... \right) \tag{21}$$

The terms above the zero order represent changes of joint distribution due to global transformation $\mathbf{U}$ and can be ignored when $\mathbf{U}$ is small. This is generally true for transformations that appear at a single registration iteration. Voxel contribution to global joint entropy, which can now be used as a voxel similarity measure is

$$S'_H(v) = \log p\left(\mathbf{i}'(v)\right), \tag{22}$$

and corresponding point similarity function is

$$S_H(\mathbf{i}) = \log p(\mathbf{i}). \tag{23}$$

Notice that point similarity $S_H(v)$ is actually estimated using the whole images, which defines probabilities $p(\mathbf{i})$ used to obtain point similarity function $S_H(\mathbf{i})$. Point similarities $S_H(v)$ then represent information of intensity pair $\mathbf{i}$ located at voxel v.

Because of using $p(\mathbf{i})$ instead of unknown $p'(\mathbf{i})$ point similarity function is suited only for small transformations $\mathbf{U}$. To comply with large transformations, similarity function $S_H(\mathbf{i})$ should be recomputed frequently, e.g. in each iteration of non-rigid registration.

Mutual information MI differs from joint entropy in marginal entropies $H(A)$ and $H(B)$, which serve as normalization factors, to make MI less dependent on image overlap and other undesirable joint distribution changes. Point similarity measure can be derived from MI using

the same approach as shown for entropy $H(A, B)$. The obtained point similarity function $S_{MI}(\mathbf{i})$ has the following form:

$$S_{MI}(\mathbf{i}) = \log \frac{p(\mathbf{i})}{p(i_A) \cdot p(i_B)}. \tag{24}$$

Point similarity measures can also be derived from other global similarity measures. Point similarity measure derived from energy similarity measure $S_G = \sum_{\mathbf{i}} p^2(\mathbf{i})$ [5],using the same approach again, differs from the entropy based measure only in the log function. Its similarity function $S_P(\mathbf{i})$ equals the estimated joint probability:

$$S_P(\mathbf{i}) = p(\mathbf{i}). \tag{25}$$

Point similarity measures can also be defined by using conditional probabilities [23, 11]:

$$S_{PC}(\mathbf{i}) = p(i_A|i_B). \tag{26}$$

$$S_{HC}(\mathbf{i}) = \log p(i_A|i_B). \tag{27}$$

All these point similarity measures are based on global joint intensity distribution, which indicates relation between intensities of both images.

2.2. Relation between point similarity measures and joint distribution

As shown earlier multi-modal point similarity measures depend on joint distribution. Joint intensity distribution is a quantitative representation of the intensity relationship at a certain image alignment, and changes during the registration [34]. Only the joint distribution of correctly matched images correspond to a real intensity relationship. However, a real intensity relationship is not known until the images are correctly registered. To better understand multi-modal point similarity measures, let us illustrate how the degree of image misalignment reflects in joint intensity distribution.

Imagine we have two simple images A and B representing the same object, see Fig. 2. Let each image consist of only two intensity values, i_{1A} and i_{2A} for image A, and i_{1B} and i_{2B} for image B, where i_{1A} corresponds to i_{1B} and i_{2A} corresponds to i_{2B}. When images are correctly registered, the joint distribution consists of only two extrema, at intensity pairs $[i_{1A}, i_{1B}]$ and $[i_{2A}, i_{2B}]$, because the intensity regions perfectly overlap. Let us call these intensity pairs true intensity pairs, as they correspond to correctly matched image regions. When images do not overlap exactly, additional 'false' intensity pairs appear, in our case $[i_{1A}, i_{2B}]$ and/or $[i_{2A}, i_{1B}]$. Probabilities of these true and false intensity pairs depend on the size of overlapping regions.

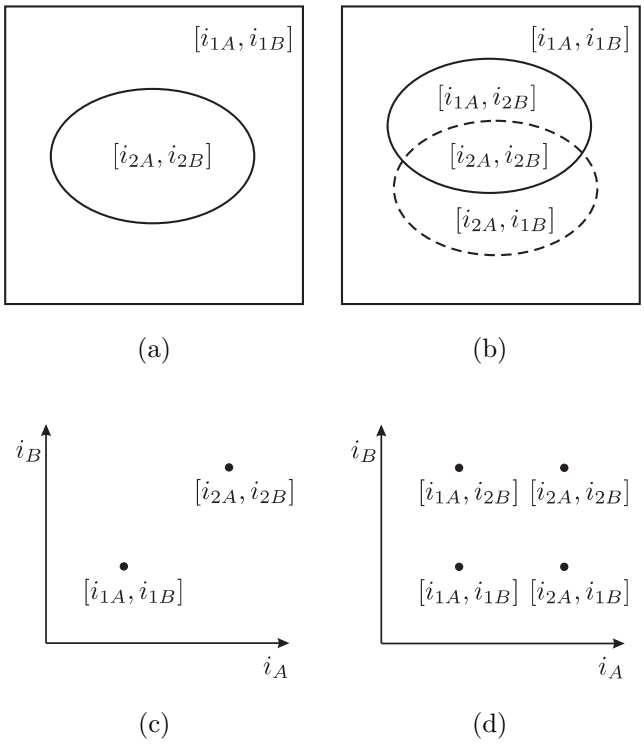
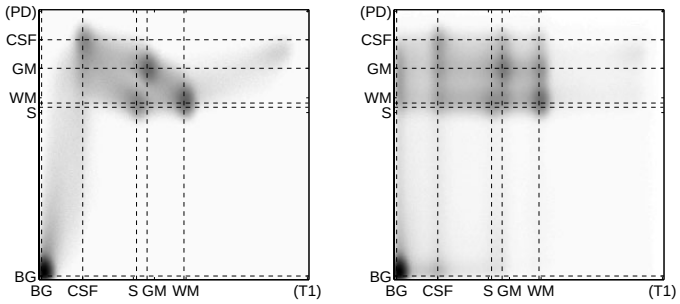


FIG. 2 Joint distribution (c) of registered images (a) and joint distribution (d) of misregistered images (b).

Real images are subjected to noise and intensity distortion, and so each region overlap is represented by numerous intensity pairs that form an intensity class. Therefore, true and false intensity pairs become true and false intensity classes. Furthermore, real images consist also of some intensities that do not belong to any of the mentioned intensity classes. In case of 3D anatomical medical images most of them appear on transitions between different tissues and belong to partial volume voxels. In joint distribution they are positioned inside a rectangle defined by corresponding intensity classes, see examples in Fig. 3.

The similarity function should model intensity dependence of correctly registered images and should have large values only for intensity pairs that belong to true classes, while similarity values for false classes should be minimal. Furthermore, intensity pairs from true classes should all produce comparable similarity. Unfortunately similarity functions differ from such ideal form. A higher level of image misalignment reflects in higher estimated probabilities for false classes, while probabilities for true classes decrease. Described similarity functions do not



(a) (b)

FIG. 3 Examples of joint distributions for MRI T1 and PD medical images of head: registered images (a) and not registered images (b). Dashed lines correspond to mean intensity values of tissue types: cerebrospinal fluid (CSF), gray matter (GM), white matter (WM), surrounding tissues (S), and background (BG). Partial volume voxels are positioned inside rectangles defined by these lines.

discriminate between true and false classes and are directly related to estimated probabilities. Thus, if the level of misalignment is too high, similarities for false classes become larger than similarities for true classes, and non-rigid registration in such case does not converge to the correct image match. Moreover, similarities of intensity pairs also depend on the amount of tissues that form these intensities. This may result in differences in their registration speed and registration correctness. To solve these problems we develop a point similarity measure, which combines registration with segmentation.

2.3. Segmentation based point similarity measure

Image registration does not directly register biological tissues, it registers their image representation. Each tissue type is represented by a class of intensity pairs, which are treated independently. If their close relations within the same tissue type are taken into account, an improvement in the registration can be expected. Furthermore, such grouping of intensity pairs can take advantage of knowledge about behaviour of intensity distributions, described in the previous section. Note that during registration process we do not know the correct joint distribution, i.e. joint distribution of registered images.

By modeling the intensity classes, a point similarity function can be estimated as a probability of an intensity pair belonging to one of the true classes. Such an estimation improves two similarity function properties. Firstly, it equalizes similarities among correctly matched tissues.

Namely, probability that an intensity pair belongs to a certain true class is not related to the amount of corresponding tissue. Secondly, such a similarity measure reduces similarities for incorrectly matched regions, as knowledge of intensity distribution can be used to distinguish between true and false classes, see Section 2.2 and Figs. 2 and 3.

This is the basic idea of segmentation based point similarity measure, which models joint intensity distribution as a sum of intensity classes obtained by segmentation. We define the segmentation based point similarity function as

$$S_S(\mathbf{i}) = p(\mathcal{C}_T|\mathbf{i}) = \sum_m p(C_m|\mathbf{i})p(\mathcal{C}_T|C_m), \tag{28}$$

where C_m denotes intensity class, which represents certain tissue type pair, and $\mathcal{C}_T$ is a set of all true intensity classes. Remember that true classes are classes which correspond to the same tissue in both images. The first factor of (28) therefore represents probability of m-th class for a given intensity pair $\mathbf{i}$. The second factor is a probability that certain class C_m is a true class and therefore belongs to $\mathcal{C}_T$.

The first step towards segmentation based similarity measure is modeling of joint distribution, which consists of several classes, each of them representing a different tissue type pair. Joint distribution $p(\mathbf{i})$ is thus a weighted sum of all class distributions $p(\mathbf{i}|C_m)$.

$$p(\mathbf{i}) = \sum_m p(\mathbf{i}, C_m) = \sum_m p(\mathbf{i}|C_m)p(C_m). \tag{29}$$

The probabilities can be obtained by segmenting the images into the tissue types and then estimating the distribution of each tissue type pair by computing its joint distribution. This approach requires segmentation of images, which is also not an easy task. Various image segmentation methods can be used, for review see [7, 28]. Note that image segmentation is in close relation with modeling intensity distribution. Not only that joint distribution can be modeled by using segmentation, the segmentation is also often based on modeling intensity distribution.

We propose another approach, which models joint distribution directly and is equivalent to simultaneous segmentation of both images. Because images are not yet registered the number of intensity classes is not known in advance and precise modeling is difficult, as most standard approaches cannot be directly used. Our approach automatically detects how many intensity classes exist and is computationally efficient, as it models only required single-tissue intensities, excluding mixtures that appear in partial volume voxels. Assuming Gaussian intensity distribution within each tissue type, each class C_m can be modeled by a 2D Gaussian function with

mean value μ_m and covariance matrix $\mathbf{\Sigma}_m$. Low intensities, e.g. background, which are approximately Rayleigh distributed can be modeled by a Gaussian model sufficiently well, because the significant difference appears only at lowest intensities, where probability of other tissues is low. Let us assume that classes are far enough from each other to achieve dominance of class C_m in its neighborhood $\mathcal{O}_m$, such that contributions of all other classes can be neglected. Classes that are merging can be modeled together as a single class and separated in later registration steps, when their overlap decreases. Number of classes M, their mean values μ_m and joint intensity distribution maxima a_m can be estimated by an exhaustive search for maxima in joint intensity distribution. When maximum is found, its position is used as a class mean value μ, while the value itself is used as an amplitude of probability $p(\mathbf{i}, C_m)$. Probabilities of intensity pairs in $\mathcal{O}_m$ can be approximated by

$$p(\mathbf{i})|_{\mathbf{i}\in\mathcal{O}_m} \approx p(\mathbf{i}, C_m) = a_m \exp\left(-\frac{1}{2}(\mathbf{i}-\mu_m)^T \mathbf{\Sigma}_m^{-1}(\mathbf{i}-\mu_m)\right) \quad ; \quad m = 1 \ldots M. \tag{30}$$

By taking a logarithm of (30) we get

$$2\ln\left(\frac{a_m}{p(\mathbf{i})}\right) = (\mathbf{i}-\mu_m)^T \mathbf{\Sigma}_m^{-1}(\mathbf{i}-\mu_m) \quad ; \quad \mathbf{i} \in \mathcal{O}_m \tag{31}$$

$$\mathbf{\Sigma}_m^{-1} = \begin{bmatrix} u_{11} & u_{12} \\ u_{12} & u_{22} \end{bmatrix} \tag{32}$$

$$2\ln\left(\frac{a_m}{p(\mathbf{i})}\right) = u_{11}(i_A-\mu_{mA})^2 + 2u_{12}(i_A-\mu_{mA})(i_B-\mu_{mB}) + u_{22}(i_B-\mu_{mB})^2 \quad ; \quad \mathbf{i} \in \mathcal{O}_m, \tag{33}$$

which can be solved for $\mathbf{\Sigma}_m^{-1}$ using least squares method for all intensity pairs $\mathbf{i}$ in the neighborhood $\mathcal{O}_m$. The covariance matrices $\mathbf{\Sigma}_m$ can then be used to estimate the class a priori probabilities $p(C_m)$.

$$\begin{aligned} p(C_m) &= \int p(\mathbf{i}, C_m) d\mathbf{i} = \\ &= \int a_m \exp\left(-\frac{1}{2}(\mathbf{i}-\mu_m)^T \mathbf{\Sigma}_m^{-1}(\mathbf{i}-\mu_m)\right) d\mathbf{i} \\ &= a_m 2\pi|\mathbf{\Sigma}_m| \quad ; \quad m = 1..M \end{aligned} \tag{34}$$

Theoretically the sum of all a priori probabilities $p(C_m)$, $m = 1...M$, should be 1. In reality this is seldom the case even if all class parameters are estimated absolutely correct, due to the fact that some intensity pairs with low joint probabilities do not belong to any of the estimated classes. The majority of such intensity pairs represents partial volume (PV) voxels. The correct position of PV voxels in joint distribution is hard to predict, as they can be positioned anywhere in the rectangle defined by the true classes, see Fig. 3. Currently we

are not dealing with partial volume voxels explicitly. We simply model them with additional class C_0 with uniform distribution:

$$p(\mathbf{i}, C_0) = \varepsilon. \tag{35}$$

The selection of ε is not critical and we set it to $1/N$ where N is a total number of overlapping voxels.

A posterior probability $p(C_m|\mathbf{i})$ of class C_m, which shows the chance that certain intensity pair $\mathbf{i}$ belongs to particular class C_m, is according to Bayes rule

$$p(C_m|\mathbf{i}) = \frac{p(\mathbf{i}, C_m)}{\sum_{l=0}^{M} p(\mathbf{i}, C_l)}. \tag{36}$$

This concludes the modeling of joint intensity distribution. Note that other modeling or segmentation approaches can be used as well. The described method is used because of its low computational cost at sufficient accuracy. The obtained model depends on correctness of match and improves during the registration. Thus, image registration and image segmentation/modeling are actually performed simultaneously.

When joint distribution model is obtained, probabilities $p(\mathcal{C}_T|C_m)$ that certain class C_m is a true class must be estimated in order to determine point similarity function (28). The set $\mathcal{C}_T$ includes all the true classes. We cannot know exactly which classes are true classes, but using knowledge of joint distributions, it is possible to estimate this probability for each of the classes. Let us assume that each tissue type has a unique intensity representation with mean value μ. Therefore, among all maxima positioned at the same intensity of image A (or image B) only one can belong to the set of true classes $\mathcal{C}_T$. Let a set of classes $\mathcal{C}_{\mu A}$ consist of all classes with the same mean value μ_A (we allow a difference of one standard deviation). Then it is expected that classes $C_m \in \mathcal{C}_{\mu A}$ with higher probabilities $p(C_m)$ are more likely to be true classes. Therefore, we can estimate probability $p_A(\mathcal{C}_T|C_m)$ of class C_m being a true class according to image A as follows:

$$p_A(\mathcal{C}_T|C_m) = \frac{p(C_m)}{\sum_{C_l \in \mathcal{C}_{\mu A}} p(C_l)}; \quad C_m \in \mathcal{C}_{\mu A}. \tag{37}$$

Such a probability estimation of a certain class being a true class is not sufficiently consistent. For example, when one of the tissues dominates, all classes with μ_B that correspond to that tissue could have the highest probability $p_A(\mathcal{C}_T|C_m)$, although it is expected that only one of them is a true class. This problem can be resolved by using probability $p_B(\mathcal{C}_T|C_m)$ of class

C_m being a true class according to image B. Estimation of this probability is equivalent to the estimation of $p_A(\mathcal{C}_T|C_m)$:

$$p_B(\mathcal{C}_T|C_m) = \frac{p(C_m)}{\sum\limits_{C_l \in \mathcal{C}_{\mu B}} p(C_l)}; \quad C_m \in \mathcal{C}_{\mu B}, \tag{38}$$

where a set $\mathcal{C}_{\mu B}$ comprises all the classes with the same μ_B. Class C_m can be assumed to be a true class only if it is a true class according to image A as well as to image B. The estimates $p_A(\mathcal{C}_T|C_m)$ and $p_B(\mathcal{C}_T|C_m)$ are based on different observations so they can be considered independent. Therefore, the final estimate of probability that class C_m is a true class can be obtained as a product of both probabilities:

$$p(\mathcal{C}_T|C_m) = p_A(\mathcal{C}_T|C_m) \cdot p_B(\mathcal{C}_T|C_m). \tag{39}$$

If two or more tissue types have the same intensity representation μ_A or μ_B our presumption is incorrect. Let us suppose there are ω tissue types with the same μ_A or μ_B and therefore, there should also be ω corresponding true classes. Our estimated probabilities $p_A(\mathcal{C}_T|C_m)$ or $p_B(\mathcal{C}_S|C_m)$ of these classes are reduced, in general ω times. Nevertheless, ratios between probabilities $p(\mathcal{C}_T|C_m)$ of these true classes and their competitive false classes remain in proportion and so the registration should still tend to correctly minimize the false classes, although it may require more registration iterations. Of course, tissues with the same μ_A or μ_B cannot be distinguished.

Once the required probabilities are estimated (i.e. $p(C_m|\mathbf{i})$ and $p(\mathcal{C}_T|C_m)$), segmentation based point similarity function $S_S(\mathbf{i})$ can be computed using (28).

Note that segmentation based point similarity measure S_S requires that the image data form intensity classes, which can be modeled by Gaussian functions. This is a reasonable assumption and not need to be exactly true in practice. Due to such modeling of joint distribution measure S_S does not model partial volume voxels. Matching of partial volume voxels is difficult as their intensity relations are not known. Linear intensity relation is often suggested, but due to multimodality this may not be the case. Measure S_S does not assume any relation between partial volume voxel intensities. It only matches pure tissues and when a correct spatial deformation model is used, partial volume voxels are expected to match correctly as well.

2.4. Multi-modal point similarity measures without segmentation

In some cases image intensity distribution does not form intensity classes. Intensities of medical images form classes in case of 3D anatomical imaging techniques, but in other cases

classes may not be formed. Furthermore, even if classes exist they may be hard to model when images include large intensity variations of tissue types, e.g. intensity inhomogeneity. Functional techniques do not form intensity classes at all. Matching of such images using previously described segmentation based measure S_S is not expected to give good results. However, in any case, some intensity pairs represent correct match and some other intensity pairs represent incorrect match. Every intensity of source image tends to match some intensities of target image, and vice versa. Therefore, each intensity pair can be treated as its own intensity class $C_{\mathbf{i}}$ with probability $p(C_{\mathbf{i}}) = p(\mathbf{i})$, mean value $\mu = \mathbf{i}$ and $p(C_{\mathbf{i}}|\mathbf{i}) = \delta(\mathbf{i})$, where $\delta(\mathbf{i})$ is a Dirac's delta function. If these classes are used the same way as in segmentation based similarity measure, a new point similarity measure S_U is obtained. It can be estimated as follows:

$$p_A(\mathcal{C}_T|\mathbf{i}) = \frac{p(\mathbf{i})}{p(i_A)} = p(i_B|i_A), \quad p_B(\mathcal{C}_T|\mathbf{i}) = \frac{p(\mathbf{i})}{p(i_B)} = p(i_A|i_B), \tag{40}$$

$$S_U(\mathbf{i}) = p(\mathcal{C}_T|\mathbf{i}) = p_A(\mathcal{C}_T|\mathbf{i}) \cdot p_B(\mathcal{C}_T|\mathbf{i}) = p(i_B|i_A) \cdot p(i_A|i_B) = \frac{p(\mathbf{i})^2}{p(i_A) \cdot p(i_B)}. \tag{41}$$

Here, $\mathcal{C}_T$ is a set of all intensities that represent correct match, while $p_A(\mathcal{C}_T|\mathbf{i})$ and $p_B(\mathcal{C}_T|\mathbf{i})$ are probabilities that intensity pair $\mathbf{i}$ represents correct match according to images A and B, respectively, which are actually conditional probabilities. Probability $p_B(\mathcal{C}_T|\mathbf{i})$ is equivalent to point similarity $S_{PC}(\mathbf{i})$, see (26). The difference between measures S_U and S_{PC} is therefore in conditional probability $p(i_B|i_A)$, which is expected to improve registration of non dominant tissues. Although grouping intensities into classes is not used, measure S_U still incorporates some knowledge about joint distributions, adopted from segmentation based approach.

It cannot be expected that a certain intensity in one image corresponds to only one intensity in the other image, in the same way as it can happen for mean intensity values of tissue types (μ_A and μ_B) in case of a segmentation based measure. Each intensity in one image corresponds to several intensities in the other image. Estimated probabilities $p(\mathcal{C}_T|\mathbf{i})$ depend on the amount of noise, which can differ according to tissue type. To reduce the differences between similarities of correctly registered intensity pairs, a logarithmic function can be applied in the same way as in case of entropy based point similarity measure $S_H(\mathbf{i})$, to depend on uncertainty rather than probability,

$$S_{UH}(\mathbf{i}) = \log p(\mathcal{C}_T|\mathbf{i}) = \log\left(p\left(i_B|i_A\right) p\left(i_A|i_B\right)\right) = \log \frac{p(\mathbf{i})^2}{p(i_A) \cdot p(i_B)}. \tag{42}$$

The obtained measure S_{UH} is related to previously described measures S_H and S_{MI}, such that

$$S_{UH}(\mathbf{i}) = \log p(\mathbf{i}) + \log \frac{p(\mathbf{i})}{p(i_A) \cdot p(i_B)} = S_H(\mathbf{i}) + S_{MI}(\mathbf{i}). \tag{43}$$

2.5. Selection of point similarity measures

So far we have described different methods to estimate point similarity functions. An important question is which measure to use for solving a given registration problem. It is well known that some similarity measures, specifically mutual information measures, are so general that they can be used for almost any kind of images. However, better results may be obtained by using appropriately more constrained measures. Estimation of a parameter that is known in advance is not reasonable, as the result may be biased due to an imperfect estimation. Moreover, estimation of additional parameters requires additional information, or the uncertainty of results increases. Obviously, the best results can be expected when the model fits well to the real situation.

Selection of a similarity measure should follow the same principle. Specifically, a similarity measure should be chosen according to the characteristics of the images. We have already classified similarity measures into mono-modal and multi-modal. However, additional classifications are possible [30]. Mono-modal similarity measures could be further classified according to model assumptions into $i_B = i_A$, $i_B = k \cdot i_A$ or $i_B = k \cdot i_A + n$. Similarly, multi-modal similarity measures can be classified into measures with functional intensity dependencies and those with statistical intensity dependencies. Furthermore, multi-modal similarity measures could be classified according to the model assumptions into those that model intensity classes, and those where no intensity classes are modeled.

Let us give some examples. Firstly, non-rigid registration of images taken in a single time series can often presume mono-modal intensity dependency $i_B = i_A$, and thus it is expected that best results can be obtained when using similarity measures MAD, MSD or their derivatives. Secondly, for registration of 3D anatomical images, in the case that intensity classes can be modeled, good results are expected by using $S_S(\mathbf{i})$ (28). Thirdly, when intensity classes cannot be modeled, measures $S_U(\mathbf{i})$ (41) or $S_{UH}(\mathbf{i})$ (42) are preferred. Finally, images that include functional information may use $S_U(\mathbf{i})$ or $S_{UH}(\mathbf{i})$ measures as well.

3. REGISTRATION BASED ON POINT SIMILARITY MEASURES

Point similarity measures enable estimation of similarity for any image point pair. As such they enable extreme locality and precision of non-rigid registration. These measures do not consider any relation between image points so they do not restrict image deformation in any sense, which is solely in the domain of deformation model. Consequently, image registration

can be divided into two functionally independent stages. The first one is estimation of external forces, which are based on point similarity measures. External forces (or external energy gradients) drive the registration process and are obtained from current, i.e. already obtained image alignment. The second stage is spatial deformation model, which converts external forces into actual image deformation, by considering some geometric constraints, e.g. elasticity.

A basic scheme of our registration method is shown in Fig. 4. It consists of two stages, which are iterated. The first stage is estimation of external forces $\mathbf{F}$ that drive the registration of image B to A. The second stage is spatial deformation model that defines the relationship between $\mathbf{F}$ and the deformation of B, described by a displacement field $\mathbf{U}$. Upon convergence of the method, the deformed version of B (now in register with A) is obtained by applying the displacements $\mathbf{U}$, as shown in Fig. 5.

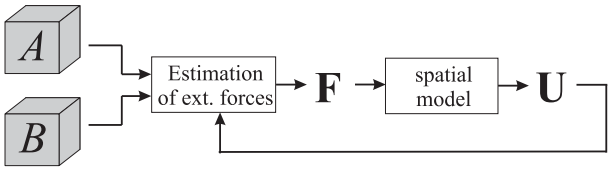


FIG. 4 Non-rigid registration scheme. The overall registration consists of two parts: estimation of external forces $\mathbf{F}$ and spatial deformation model that imposes dependencies on voxel displacements $\mathbf{U}$.

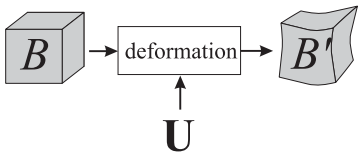


FIG. 5 Deformation of the source image B into the resulting image B' by applying displacements $\mathbf{U}$.

Force estimation approaches in general differ in two aspects: the similarity measure used to compare images in order to detect local discrepancies, and the method used to derive forces from the similarities. We have implemented two force estimation methods. The first one is a widely used force estimation, which we call forward force estimation, defined as a gradient of local similarity with respect to the reference image:

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}_F(\mathbf{x}) = \left.\frac{\partial}{\partial \boldsymbol{\varepsilon}}\right|_{\boldsymbol{\varepsilon}=0} S\left(i_A\left(\mathbf{x}+\mathbf{U}(\mathbf{x})+\boldsymbol{\varepsilon}\right), i_B\left(\mathbf{x}\right)\right). \tag{44}$$

Here, $\mathbf{x} = [x_1, x_2, x_3]^T$ denotes spatial coordinates of original undeformed images, $\mathbf{U}(\mathbf{x})$ is current point displacement obtained by registration, and $\boldsymbol{\varepsilon}$ is a displacement of a point from an already obtained configuration. Forward forces tend to move source image voxels towards their better match in the target image.

The second method, which we call consistent force estimation, is related to consistent image registration [6, 36]. In this case the external forces consist not only of forward forces $\mathbf{F}_F$ but also of reverse forces $\mathbf{F}_R$, which tend to improve matching of points in the target image according to the source image:

$$\mathbf{F}_R(\mathbf{x}) = \left.\frac{\partial}{\partial \boldsymbol{\varepsilon}}\right|_{\boldsymbol{\varepsilon}=0} S\left(i_A\left(\mathbf{x}+\mathbf{U}(\mathbf{x})\right), i_B\left(\mathbf{x}+\boldsymbol{\varepsilon}\right)\right). \tag{45}$$

The consistent external forces are then defined as:

$$\mathbf{F} = \mathbf{F}_F - \mathbf{F}_R. \tag{46}$$

In our implementation the estimation of similarity gradients in (44) and (45) follows [1]. The design of second registration stage, i.e. spatial deformation model, follows the concept of Bro-Nielsen and Gramkow [2]. The idea is that a linear model can be separated into two parts:

$$\mathbf{U}_F = k_E \mathbf{F}, \tag{47}$$

$$\mathbf{U}(\mathbf{x}) = \int \mathbf{U}_F(\mathbf{x}-\mathbf{r})\mathbf{G}(\mathbf{r})d\mathbf{r} = (\mathbf{U}_F \otimes \mathbf{G})(\mathbf{x}) = ((k_E\mathbf{U}) \otimes \mathbf{G})(\mathbf{x}). \tag{48}$$

The first part (47) is Hooke's law, which says that the point moves proportionally to the applied force $\mathbf{F}$ for displacement $\mathbf{U}_F$. The second part (48) is a spatial convolution filter and models interdependence of points. The kernel for elastic media is described in [2], but for simplicity we use Gaussian filtering. Note that $\mathbf{F}$ appears because of an imperfect match and thus the displacements $\mathbf{U}_F$ are only corrections of already obtained configuration, while $\mathbf{U}$ stands for overall displacements, measured from the initial (undeformed) image configuration.

By the principle of linearity final displacements can be calculated by summing up partial displacements,

$$\mathbf{U}^{(t)} = \mathbf{U}^{(t-1)} + \mathbf{U}_F^{(t)} \otimes \mathbf{G}, \tag{49}$$

where t denotes iteration number. This is the concept of incremental deformations [1] used to accommodate large nonlinear deformations, but because external forces $\mathbf{F}(\mathbf{x})$ are highly nonlinear function of displacements $\mathbf{U}(\mathbf{x})$ [12], the resulting deformation does not necessarily follow the selected (linear) spatial deformation model. This can be compensated by introducing

point dependence on the accumulated displacements:

$$\mathbf{U}^{(t)} = (\mathbf{U}^{(t-1)} + \mathbf{U}_F^{(t)}) \otimes \mathbf{G}. \tag{50}$$

In this way the displacements are filtered throughout the iteration process, such that old forces contribute less than later ones. The undesirable side effect of this filtering is that as the external forces go to zero, the image gradually returns back to its undeformed configuration. Thus, additional external forces are needed to sustain the deformed condition. Therefore, we simply combine the two spatial deformation models (Fig. 6):

$$\mathbf{U}^{(t)} = (\mathbf{U}^{(t-1)} + \mathbf{U}_F^{(t)} \otimes \mathbf{G}_1) \otimes \mathbf{G}_2 \tag{51}$$

The first filter ($\mathbf{G}_1$) enables large deformations and precise registration while the second one ($\mathbf{G}_2$) serves to improve the linearity of the results. The coefficient k_E controls registration speed and changes during the registration to allow fast convergence as well as high precision. Initially it is set to such a value that maximal force $\mathbf{F}(\mathbf{x})$ causes displacement $U_F(\mathbf{x})$ of one voxel, and decreases during the registration. This approach normalizes the forces so that their scaling does not affect registration results.

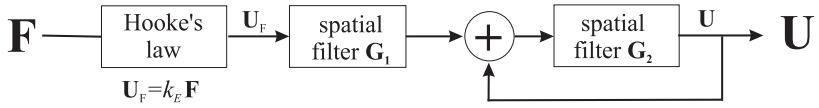


FIG. 6 Spatial deformation model.

The registration uses a multi-resolution approach, which starts at lower image resolutions that remove image differences with large spatial extent, and continues with higher resolutions that remove more and more detailed image discrepancies. To maintain the same intensity resolution in all levels of multi-resolution approach, we use Parzen window estimation [26] for obtaining the intensity distributions. The registration stops after a predefined number of iterations.

4. RESULTS

A reasonable approach to compare global similarity measures is to observe similarities according to the applied image transformation, e.g. displacement of whole image, and evaluate some properties of such similarity function, e.g. number of local extrema, smoothness, position

of global extremum, capture range, etc [29]. However, comparison of point similarity measures turns out to be more problematic, as any transformation of such an extremely small image region makes a drastic change in region overlap.

Similarly to the comparison of global measures, point similarity measures could be compared by comparing their similarity functions. The criteria could include: number of maxima that do not represent a correct match, differences among similarities estimated for incorrectly matched tissues, differences among similarities estimated for correctly matched tissues, etc. However, comparison of these properties does not directly tell us which point similarity measure is more appropriate.

It appears that the only meaningful way to compare point similarity measures is to compare registration results obtained by using these measures. In the rest of the section we focus only on the multi-modal measures. We have compared them using our non-rigid registration system, described in Section 3. To show performance of point similarity measures under different registration configurations, we have compared them using both force estimation methods. We have used four resolution levels, each consisting of 10 registration iterations and used a Gaussian filters with $\sigma = 3$ for $\mathbf{G}_1$ and $\mathbf{G}_2$.

We have used simulated and real images to compare multi-modal point similarity measures. First we have used MRI T1 and PD Brainweb simulated images [18] with $1 \times 1 \times 1$ mm voxel size. The tests were performed using images of whole head and images of brain only, both of them in three different image qualities: normal images, 9% noisy images and 40% intensity inhomogeneous images. In total 6 simulated image pairs were used. The same experiment was performed using real MRI images. For that purpose we used MRI T1 image as target with voxel size $0.86 \times 0.86 \times 0.99$ mm, while the MRI PD source image had $0.98 \times 0.98 \times 1.1$ mm voxel size. The images were previously corrected for intensity inhomogeneity using an approach based on information minimization [22]. This comparison of similarity measures is more demanding as the quality of real images is worse than the quality of simulated images. Namely, target image (MRI-T1) had poor intensity resolution, while source image (MRI-PD) included large image artifacts (intensity inhomogeneity and wrap arround artifacts).

Our comparison of multi-modal point similarity measures is based on recovering a synthetic deformation, see Fig. 7. The evaluation procedure makes use of two images, A and B, that are already in register, e.g., acquired at the same time but using different acquisition modalities or protocols. Image B is deformed using some known transformation U_0 to obtain image B', which is then registered to image A. Ideally, the obtained transformation U_R should equal the

known inverse transformation U_0^{-1}. The quality of the registration can thus be measured by the RMS (root-mean-squared) residual difference e_{RMS} between displacements $U_0^{-1}(\mathbf{x})$ and $U_R(\mathbf{x})$:

$$e_{RMS} = \sqrt{\frac{1}{N_\Omega} \sum_{\mathbf{x} \in \Omega} \left(U_0^{-1}(\mathbf{x}) - U_R(\mathbf{x}) \right)^2} \quad , \quad \mathbf{x} = [x_1, x_2, x_3]^T. \tag{52}$$

Non-object voxels are excluded from the evaluation through the specification of the object domain Ω (in practice, a binary mask defined over the image), where N_Ω denotes the number of voxels in Ω. We used a synthetic deformation generated as a sum of Gaussian functions.

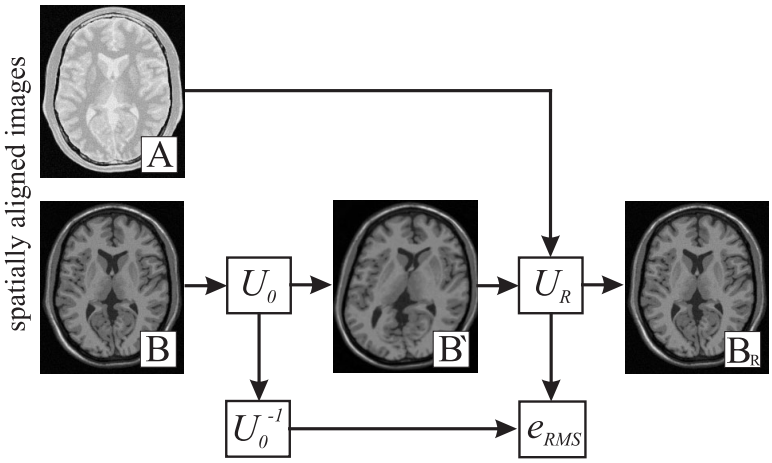


FIG. 7 Evaluation scheme for non-rigid registration, based on synthetically deformed images.

Specifically, six functions with standard deviation ranging between 15 and 60 mm were used, resulting in initial displacement error $e_{rms} = 6.90$ mm for simulated Brainweb images, and $e_{rms} = 14.15$ mm for real images. The difference between real and simulated images arose due to different position and orientation of head. We cannot argue that Gaussian deformation can be used to model actual deformations in clinical multi-modal applications. However, due to functional independence of force estimation stage and deformation model used, the results obtained that way are appropriate for comparing force estimation methods and their fundamental part - similarity measures. While the spatial deformation model may favor some type of deformation, ordering of different external force estimation approaches is still preserved.

The system was implemented in Borland C++ Builder and runs under MS Windows platforms. Registration of each image pair takes about 20 minutes on a 500 MHz Pentium III workstation with 1GB of RAM. The results of registering simulated images are tabulated in Table 1. For all compared similarity measures sensitivity to noise is low while sensitivity to intensity inhomogeneity (shading) is relatively high. The most sensitive to intensity inhomo-

TABLE 1
Comparison results of multi-modal point similarity measures using simulated Brainweb images. All results are in millimeters and represent residual errors after registration of whole head images or brain images.

Images of whole head, forward force estimation:

measure	normal		9% noise		40% shading	
	e_{RMS}	e_{max}	e_{RMS}	e_{max}	e_{RMS}	e_{max}
S_P	5.35	14.21	5.08	14.51	6.15	15.32
S_H	0.89	5.39	0.99	5.77	22.70	44,09
S_{MI}	1.09	6.83	1.23	7.29	3.00	10.01
S_{PC}	10.20	22.04	15.69	26.21	22.63	43.64
S_{HC}	0.86	5.35	1.03	6.14	23.02	44.56
S_S	0.68	2.44	0.73	3.09	3.72	10.71
S_U	2.78	10.42	3.17	11.41	5.97	15.01
S_{UH}	0.83	4.90	1.09	6.36	2.68	9.18

Images of whole head, consistent force estimation:

measure	normal		9% noise		40% shading	
	e_{RMS}	e_{max}	e_{RMS}	e_{max}	e_{RMS}	e_{max}
S_P	3.67	12.32	3.48	12.18	4.65	13.54
S_H	0.80	4.50	0.87	4.93	7.47	24.13
S_{MI}	0.83	5.04	0.99	6.05	2.39	8.09
S_{PC}	1.05	6.87	1.47	8.35	3.11	10.07
S_{HC}	0.71	4.23	0.91	5.43	15.69	37.85
S_S	0.51	1.91	0.55	2.08	4.02	11.12
S_U	2.00	9.22	2.00	9.31	4.55	13.21
S_{UH}	0.72	4.08	0.79	4.59	2.59	9.25

Images of brain, forward force estimation:

measure	normal		9% noise		40% shading	
	e_{RMS}	e_{max}	e_{RMS}	e_{max}	e_{RMS}	e_{max}
S_P	5.20	13.24	5.72	14.78	7.03	16.59
S_H	0.73	3.99	0.99	5.88	2.14	8.95
S_{MI}	1.65	9.17	1.79	9.32	2.42	10.18
S_{PC}	0.72	3.48	0.95	3.50	3.01	9.98
S_{HC}	1.29	7.69	1.54	8.50	2.24	9.61
S_S	0.55	2.19	0.77	2.58	3.96	11.82
S_U	1.57	4.25	1.98	4.49	8.04	17.22
S_{UH}	0.56	2.56	0.68	3.23	1.54	5.96

Images of brain, consistent force estimation:

measure	normal		9% noise		40% shading	
	e_{RMS}	e_{max}	e_{RMS}	e_{max}	e_{RMS}	e_{max}
S_P	3.35	12.18	3.36	12.21	3.42	12.29
S_H	0.50	2.50	0.65	3.36	2.09	9.45
S_{MI}	0.95	6.07	1.11	6.89	1.92	9.48
S_{PC}	0.98	6.14	1.10	6.39	2.28	9.35
S_{HC}	0.65	3.40	0.79	4.33	1.91	8.94
S_S	0.48	1.98	0.59	2.19	3.22	10.14
S_U	0.63	2.49	0.84	2.77	7.72	18.65
S_{UH}	0.43	2.10	0.49	2.23	1.77	7.98

TABLE 2
Comparison results of multi-modal point similarity measures using simulated deformation and real MRI-PD/T1 images of human head for forward force estimation ($\mathbf{F} = \mathbf{F}_F$) and consistent force estimation ($\mathbf{F} = \mathbf{F}_F - \mathbf{F}_R$). All results are in millimeters.

measure	$\mathbf{F} = \mathbf{F}_F$		$\mathbf{F} = \mathbf{F}_F - \mathbf{F}_R$	
	e_{RMS}	e_{max}	e_{RMS}	e_{max}
S_P	10.17	22.95	7.11	21.17
S_H	2.67	13.80	2.07	10.91
S_{MI}	2.17	12.23	2.11	10.99
S_{PC}	11.41	25.68	2.36	11.54
S_{HC}	2.67	13.79	2.08	9.87
S_S	1.56	6.56	1.66	7.33
S_U	8.04	20.21	5.44	18.80
S_{UH}	2.32	12.50	1.83	10.22

geneity was registration of whole head images with forward force estimation, where registration was successful only when using measures S_{MI}, S_S and S_{UH}. Note that the level of intensity inhomogeneity was higher than is expected for real data. The problem appears due to intensities of tissues surrounding brain, which are similar to intensities of some brain tissue in the source image, and some other tissue in the target image. In such a situation registration can deteriorate, especially when it is based on measures S_{PC}, S_U and S_P, and in more extreme cases also measures S_H and S_{HC}. Shading has smaller influence on e_{RMS} for images of brain, as they do not include surrounding tissues, and consequently, class overlap at intensity shading is lower than in the case of images of whole head, where surrounding tissues form additional intensity classes. In general, when images are not subject to intensity inhomogeneity, the best results are mostly obtained by using measure S_S, in other cases the best results are mostly obtained by measure S_{UH}. When images include only intensity classes that are easy to distinguish, e.g. in case of registering brain images, almost every measure gives good results and differences between them are less apparent. The results also show that consistent force estimation in general gives better results than forward force estimation.

Results obtained for registering real images, which are tabulated in Table 2, are not as good as results obtained for the simulated images. There are two possible reasons. First, quality of real images is worse that quality of simulated images, and second, worse results may be caused by higher initial misalignment. However, comparison of results obtained from real images suggest similar conclusions as we made for simulated images. Segmentation based measure S_S performed best, measures S_{MI} and S_{UH} were also good, especially for consistent

force estimation. On the other hand measures S_P, S_U and eventually S_{PC} are shown to be less appropriate for such registration tasks.

The quality of point similarity measures can be observed also by comparing their similarity functions. There are two issues: firstly the ratio between similarity for correct and incorrect match, and secondly the ratio between similarity for different correctly matched intensity pairs. Examples of point similarity functions are shown in Fig. 8. All similarity functions are estimated from the same unregistered images, so that joint intensity distribution (which equals S_P) consists of true intensity classes, which represent correctly matched tissues, as well as false intensity classes, which represent incorrect match. Ideally, similarity should be high and equal for all intensity pairs that represent correct matching (true classes) and low for all other intensity pairs (including false classes).

Examples of point similarity functions show that measures S_P, S_{PC} and S_U highly depend on the amount of different tissue types. In our case background dominates over the tissues and thus registration of tissue intensities is much slower than registration of background, which results in worse registration results. A log function, which is used for measures S_H, S_{MI}, S_{HC} and S_{UH}, makes similarities for different correctly registered tissues less different. However, similarities for incorrectly registered parts become more similar to those for correct match as well. This becomes problematic especially in case of measure S_{HC}. Similarity functions for S_H, S_{MI} and S_{UH} are related (43). S_H is directly related to joint intensity distribution p and produces high similarity for intensity pairs with high probability $p(\mathbf{i})$, e.g. for true classes. On the other hand measure S_{MI} provides good discrimination between true and false classes, while similarity is higher for less dominant true classes and relatively low for dominant ones, e.g. for background. This is disadvantageous when high tissue misalignment exists, which is common for initial registration steps. Measure S_{UH} improves that property, and provides high similarities for all true classes, while discrimination between true and false classes is still good. Similarity function for measure S_S differs from other measures, as it models the joint distribution as a sum of intensity classes, while other intensity pairs, e.g. partial volume voxels, are not modeled. In comparison to the other point similarity measures it better distinguishes between true and false classes.

When images do not form intensity classes or intensity classes cannot be accurately modeled, measure S_S cannot be expected to give good results. Other measures are still applicable. Measure S_{UH} seems to be the most appropriate for use in such situations, as in general it gave best results on MRI images. For illustrative purposes only, we show registration of PET-CT

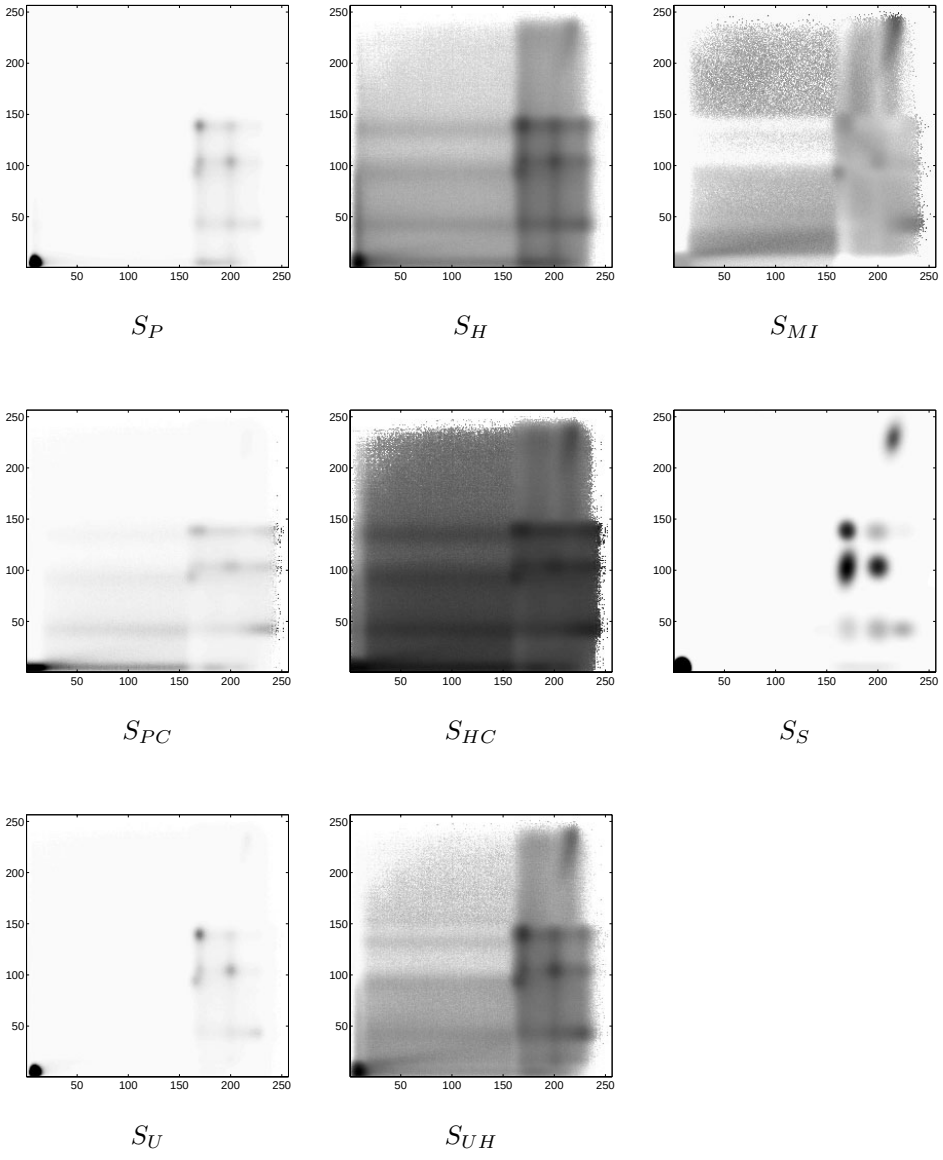


FIG. 8 Examples of similarity functions of multi-modal point similarity measures for registering MRI-T1 and MRI-PD images of head. Darker color means higher similarity. Ideally, similarity should be high and equal for all intensity pairs that represent correct matching (true classes) and low for all other intensity pairs (including false classes). Evidently S_S is better than other similarity measures.

thoracic images, see Fig. 9. Joint intensity distribution of registered images shown in Fig. 10 confirms that no intensity classes are formed.

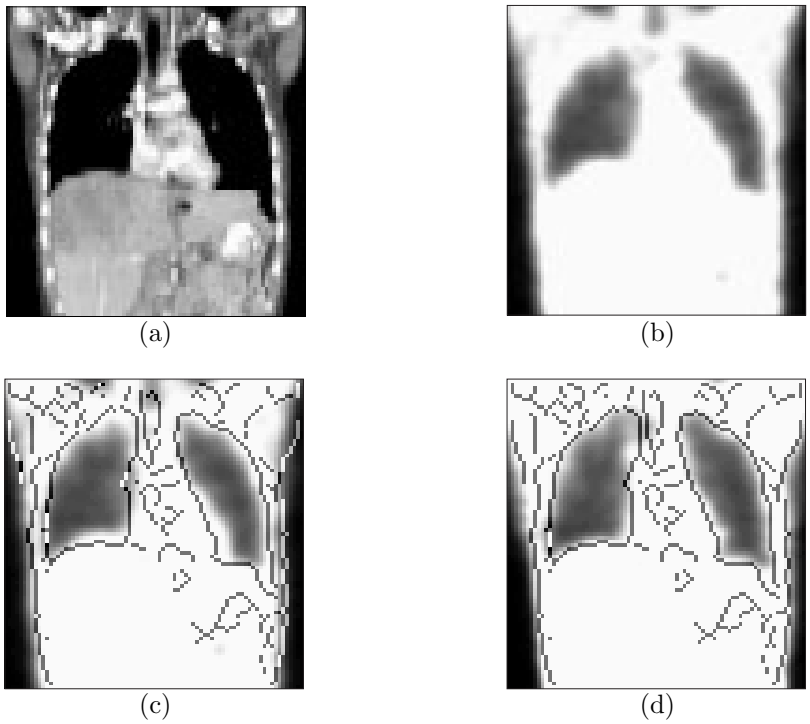


FIG. 9 PET-CT registration of thoracic images: CT image (a), PET transmission image (b), rigidly registered (c) and non-rigidly registered (d) PET transmission image overlayed with contours obtained using Canny edge detector on CT image. Note the improvement of match obtained by non-rigid registration, which is most visible in the lung section.

When joint intensity distribution forms the classes, registration based on point similarity measure S_S gives segmented images as a by-product. Segmentation and registration are actually performed simultaneously. If each intensity pair in the images is classified to its most probable intensity class, the image gets divided into multiple regions, where each region belongs to one of the intensity classes. In general images consist of true and false intensity classes. Both of them are present when images are not correctly registered, see Fig. 11 (left). Registration reduces the amount of tissues belonging to the false classes and when images are correctly registered only the true classes remain, see Fig. 11 (right). Each true class then represents a certain tissue type. The quality of segmentation gradually increases during the registration. Performing registration and segmentation simultaneously may improve registration (as it is evident from

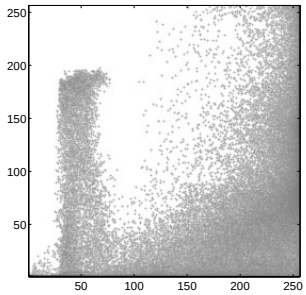


FIG. 10 Joint intensity distribution of registered CT/PET-transmission images.

the results), as well as segmentation results.

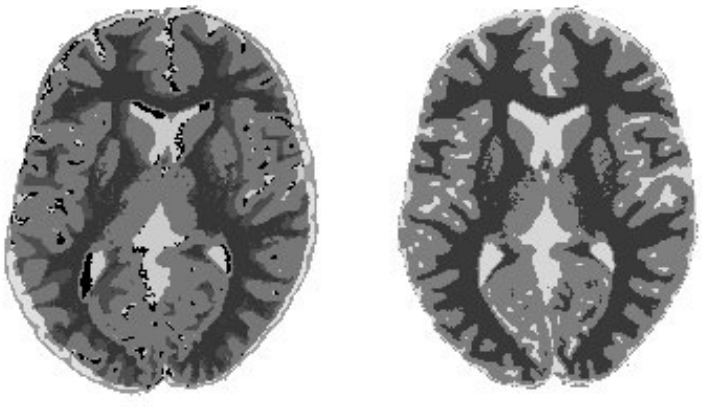


FIG. 11 Result of the segmentation for misregistered (left) and registered images (right). Each class is represented by different gray value.

5. DISCUSSION AND CONCLUSIONS

In this paper we described a new class of similarity measures for estimating local image correspondences. They work on individual voxels, but consider information derived from the whole images. Several point similarity measures were compared using one synthetic image set with different noise and shading, and one real MRI data set. The comparison was based on recovering synthetic deformation. All compared measures are based on joint intensity distribution. When intensity distribution can be modeled by intensity classes, which is related to segmentation of the images, the best registration was achieved by the so called segmentation

based point similarity measure S_S. When intensity distribution could not be modeled, the point similarity measure S_{UH} is recommended. This was demonstrated by registering a PET transmission image to a CT image.

There are some more issues that should be considered. Although point similarity measures are applied locally, the relation between image intensities is modeled globally. That allows more reliable estimation of complex multi-modal intensity relationships. However, local estimation can be also advantageous. Local properties are less sensitive to variations of intensity relation across the images, which explains why in some cases even some mono-modal measures can be used to compare multi-modal data [39]. Still, due to better estimation of intensity correspondence, global estimation is usually more reliable and misregistration is less likely to occur. The results in this paper show that point similarity measures are sensitive to a high variation of intensity relations across the image, as are all global measures, including mutual information measures. Fortunately, such image variations can be removed, e.g. by using inhomogeneity correction.

Estimation of point similarities can be improved significantly by using prior information. When joint probability of correctly matched images is known, a similarity function can be estimated from a weighted sum of the current estimate and the prior joint intensity distribution.

Some precaution is required in selection of image processing algorithms used in the registration process. Convolution filtering, interpolation, resampling and similar processes change image intensity distribution. They make intensity distributions more scattered, which deteriorates estimation of the similarity function. Furthermore, such methods presuppose linear intensity relation, which is not necessarily correct in the case of multi-modal data. Linear interpolation of intensities is therefore undesirable as well. Hence, we use linear interpolation of similarity instead of intensity, and partial volume interpolation for estimation of joint distribution.

Finally, note that point similarity meaures can be derived from almost any (global) measure. Nevertheless, good global similarity measures do not necessarily form good point similarity measures.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Slovenian Ministry of Education, Science and Sport (Research program 1538-517), and the U.S.P.H.S. for this work under grants NS33662, LM03504, MH62100, AG15116, AG17586. We thank Dr. A.W. Toga for providing the MRI dataset.

REFERENCES

[1] R. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing*, 46:1–21, April 1989.

[2] M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. *Springer Lecture Notes in Computer Science*, 1131:267–276, 1996.

[3] L.G. Brown. A survey of image registration techniques. *ACM Comput. Surveys*, 24(4):325–376, December 1992.

[4] T. M. Buzug and J. Weese. Voxel-based similarity measures for medical image registration in radiological diagnosis and image guided surgery. *Journal of Computing and Information Technology*, 6(2):165–179, 1998.

[5] T. M. Buzug, J. Weese, C. Fassnacht, and C. Lorenz. Elastic matching based on motion vector fields obtained with a histogram based similarity measure for DSA-image correction. *Computer Assisted Radiology and Surgery*, pages 139–144, 1997.

[6] G.E. Christensen and H.J. Johnson. Consistent image registration. *IEEE Transactions on Medical Imaging*, 20(7):568–582, July 2001.

[7] L. Clarke, R. Vethuizen, M. Camacho, J. Heine, M. Vaidyanathan, L. Hall, R. Thatcher, and M. Silbiger. MRI segmentation: Methods and applications. *Magnetic Resonance Imaging*, 13(3):343–368, 1995.

[8] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information processing in medical imaging 1995*, pages 263–274. Kluwer Academic, 1995.

[9] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3D multi-modality medical image registration using feature space clustering. In N. Ayache, editor, *CVRMed*, volume 905 of *Lecture Notes in Computer Science*, pages 195–204. Springer, 1995.

[10] T. Gaens, F. Maes, D. Vandermeulen, and Suetens P. Non-rigid multimodal image registration using mutual information. In W.M. Wells, A. Colchester, and S. Delp, editors, *Proceedings of the 1st International Conference on Medical Image Computing and Computer-*

Assisted Intervention – MICCAI'98, number 1496 in Lecture Notes in Computer Science, pages 1099–1106, MIT, Cambridge, MA, USA, October 1998. Springer-Verlag.

[11] J. C. Gee, D. R. Haynor, M. Reivich, and R. Bajcsy. Finite element approach to warping of brain images. In M. H. Loew, editor, *Proc. SPIE Medical Imaging 1994: Image Processing*, volume 2167. SPIE Press, Bellingham, WA, 1994.

[12] James C. Gee. On matching brain volumes. *Pattern Recognition*, 32(1):99–111, 1999.

[13] A. Guimond, A. Roche, N. Ayache, and J. Meunier. Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections. *IEEE Transactions on Medical Imaging*, 20(1):58–69, January 2001.

[14] N. Hata, T. Dohi, S.K. Warfield, W.M. Wells, R. Kikinis, and F.A. Jolesz. Multimodality deformable registration of pre- and intraoperative images for MRI-guided brain surgery. In W.M. Wells, A. Colchester, and S. Delp, editors, *Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, number 1496 in Lecture Notes in Computer Science, pages 1067–1074, MIT, Cambridge, MA, USA, October 1998. Springer-Verlag.

[15] D. L. G. Hill and D. J. Hawkes. Across-modality registration using intensity-based cost functions. In I. Bankman, editor, *Handbook of Medical Image Processing*, pages 537–553. Academic Press, 1999.

[16] M. Holden, D. L. G. Hill, E. R. E. Dent, J. M Jarosz, T. C. S. Cox, T. Rohlfing, Goodey J., and D. J. Hawkes. Voxel similarity measures for 3-D serial MR brain image registration. *IEEE Transactions on Medical Imaging*, 19(2):94–102, February 2000.

[17] P. J. Kostelec, J. B. Weaver, and D. M. Healy Jr. Multiresolution elastic image registration. *Medical Physics*, 25(9):1593–1604, September 1998.

[18] R.K.-S. Kwan, A.C. Evans, and G. B. Pike. An extensible MRI simulator for post-processing evaluation. In *Visualization in Biomedical Computing (VBC'96)*, volume 1131 of *Lecture Notes in Computer Science*, pages 135–140. Springer-Verlag, May 1996.

[19] M.E. Leventon and W.E.L. Grimson. Multi-modal volume registration using joint intensity distributions. In W.M. Wells, A. Colchester, and S. Delp, editors, *Proceedings of the 1st*

International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'98, number 1496 in Lecture Notes in Computer Science, pages 1057–1066, MIT, Cambridge, MA, USA, October 1998. Springer-Verlag.

[20] B. Likar and F. Pernuš. Registration of serial transverse sections of muscle fibers. *Cytometry*, 37(2):93–106, 1999.

[21] B. Likar and F. Pernuš. A hierarchical approach to elastic registration based on mutual information. *Image and Vision Computing*, 19:33–44, 2001.

[22] B. Likar, M.A. Viergever, and F. Pernuš. Retrospective correction of MR intensity inhomogeneity by information minimization. *IEEE trans. on Medical Imaging*, 20(12):1398–1410, 2001.

[23] J. B. A. Maintz, H. W. Meijering, and M. A Viergever. General multimodal elastic registration based on mutual information. In K.M. Hanson, editor, *Medical Imaging 1998: Image Processing*, volume 3338 of *Proc. SPIE*, pages 144–154. SPIE Press, Bellingham, WA, 1998.

[24] J.B.A. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[25] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank. Nonrigid multimodality image registration. In M. Sonka and K.M. Hanson, editors, *Medical Imaging 2001: Image Processing*, volume 4322 of *Proc. SPIE*. SPIE Press, Bellingham, WA, 2001.

[26] E. Parzen. On the estimation of probability density function. *Annual Mathematical Statistics*, 33:1065–1076, 1962.

[27] X. Pennec, N. Ayache, and J.P. Thirion. Landmark-based registration using features identified through differential geometry. In I. Bankman, editor, *Handbook of Medical Imaging*, pages 499–513. Academic Press, September 2000.

[28] D.L. Pham, C. Xu, and J.L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337, 2000.

[29] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging*, 19(8):809–814, August 2000.

[30] A. Roche, G. Malandain, and N. Ayache. Unifying Maximum Likelihood Approaches in Medical Image Registration. *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, 11(1):71–80, 2000.

[31] G. K. Rohde, A Aldroubi, and B. M. Dawant. Adaptive free-form deformation for inter-patient medical image registration. In *The Proceedings of the SPIE Symposium on Medical Imaging 2001*, 2001.

[32] K. Rohr, H.S. Stiehl, R. Sprengel, W. Beil, T.M Buzug, J. Weese, and M.H. Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Trans. on Medical Imaging*, 20(6):526–534, 2001.

[33] D. Rueckert, C. Hayes, C. Studholme, P. Summers, M. Leach, and D.J. Hawkes. Non-rigid registration of breast MR images using mutual information. In W.M. Wells, A. Colchester, and S. Delp, editors, *Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, number 1496 in Lecture Notes in Computer Science, pages 1144–1152, MIT, Cambridge, MA, USA, October 1998. Springer-Verlag.

[34] C. Studholme, Hill. D.L.G., and D.J. Hawkes. Using voxel similarity as a measure of medical image registration. In E. Hancock, editor, *Proceedings of the British machine vision conference – BMVC'94*, pages 235–244, 1994.

[35] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86, 1999.

[36] J.-P. Thirion. Image matching as a diffusion process: An analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[37] A. W. Toga, editor. *Brain Warping*. Academic Press, San Diego, 1999.

[38] P. Viola and W. Wells III. Alignment by maximization of mutual information. In *Proceedings of the 5th International Conference on Computer Vision*, pages 16–23, 1995.

[39] J. Weese, P. Rosch, T. Netsch, T. Blaffert, and M. Quist. Gray-value based registration of CT and MR images by maximization of local correlation. In C.J. Taylor and A. Colchester, editors, *Proceedings of the 2nd International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'99*, volume 1679 of *Lecture Notes in Computer Science*, pages 656–663. Springer, 1999.