Similarity Measures for Non-Rigid Registration*

Peter Rogelj and Stanislav Kovačič

Faculty of Electrical Engineering, University of Ljubljana,
Tržaška 25, 1001 Ljubljana, Slovenia,
peter.rogelj@fe.uni-lj.si, stanek@fe.uni-lj.si

Abstract Similarity measures for non-rigid multimodal registration are required to be local in order to enable correction of small image differences, and multimodal, to allow images to be acquired using different imaging techniques. Unfortunately all commonly used multimodal similarity measures are inherently global and cannot be directly used to estimate local image properties. We have derived a local similarity measure based on joint entropy, which can operate on extremely small image regions, e.g. individual voxels. The disadvantage of using such small image regions is higher sensitivity to noise and partial volume voxels, which reduce registration speed and accuracy. To cope with these problems we support the similarity measure with image segmentation. Several experiments based on synthetic images show that simultaneous application of registration and segmentation can improve registration accuracy and reduce the required number of registration steps.

1 Introduction

Non-rigid registration is a process for maximizing a spatial image correspondence of two images within the constrains of an image transformation model, to bring the features of first image into alignment with those of the second image with possibly different content. Image correspondence is measured using similarity measures, which compares the data values at corresponding points in the images. A variety of similarity measures is suitable for images obtained using the same acquisition method [1]. The situation is quite different when the images to be non-rigidly registered are acquired using different modalities. There are two main issues to be addressed: first, the similarity measure must be capable of determining the correspondence between images of different measurement type, and second, the measure must be sensitive to local image differences. The problem is made difficult because the two properties desired of our similarity measure are contradictory.

In practice, image differences appear on a range of scales, from "global" to "local". The aim of non-rigid registration is to correct those that comply with the image transformation model. Therefore, it is desired to correct even the smallest image discrepancies, but to do that they must previously be detected.

* This work was supported by the Ministry of Science and Technology of the Republic of Slovenia (Research program 1538-517)

Therefore, similarity must be estimated from correspondingly small-sized image regions, which in extreme case become individual voxels. It is noted [2,3] that registration based on individual voxels is ill-posed if displacements are calculated separately for each image voxel. For this purpose non-rigid registration must incorporate a spatial model, e.g. elastic [4] or viscous fluid model [5], which defines the relations between individual region/voxel displacements.

1.1 Multimodal similarity measures

Multimodal similarity measures are capable of determining the correspondence between images of different measurement type, and therefore different intensity values for the same anatomical structure. The most commonly used multimodal similarity measure is mutual information, first proposed and brought to the medical imaging field from information theory by Viola and Wells [6]. When used as a similarity measure it measures the statistical dependence between the image intensities. In principle, it reveals how much one image tells us about the other image, and takes on maximum value when the images are geometrically aligned. Given two images A and B mutual information $I(A, B)$ between them is defined as

$$I\left(A, B\right) = H\left(A\right) + H\left(B\right) - H\left(A, B\right), \tag{1}$$

where$H(A)$ and $H(B)$ denote marginal entropies of A and B, and $H(A, B)$ is their joint entropy. The entropies can be calculated using well-known Shannon definition:

$$H\left(\cdot\right) = -\sum p\left(\cdot\right) \log p\left(\cdot\right), \tag{2}$$

where $p(.)$ stands for either marginal or joint probability distributions, estimated from image intensities $\mathbf{i} = [i_A, i_B]^T$. There are also other types of multimodal similarity measures, like energy similarity measure [7], and various types of generalized entropies [8]. Nevertheless, due to their statistical nature they are all inherently global in the sense that they all require relatively large image regions to achieve sufficiently high statistical significance of joint distribution. Various solutions to improve their locality were proposed, but among them only local multimodal similarity measures can operate on arbitrarily small image regions, e.g. individual voxels. We have derived a similarity measure of that type from joint entropy [9]. It was observed, that entropy can be calculated by summation in spatial domain, over all voxels v:

$$H = \sum_v -\frac{1}{N} \log p_v = \frac{1}{N} \sum_v h\left(v\right), \tag{3}$$

where p_v is the joint probability $p_{\mathbf{i}}$ for intensity pair at position v, and N is the total number of voxels in the image. Thus, the entropy, which is a global similarity measure, can be calculated from similarities of individual voxels $S^v = h(v) = -\log p_v$, which represent the "uncertainty" of intensity pair located at that position. Summation can also be performed over arbitrary smaller image

region R_r and the result is local similarity S^r of this region.

$$S^r = \frac{1}{N_r} \sum_{v \in R_r} h\,(v). \tag{4}$$

Notice that $h(v)$ is always estimated using the whole images. The local similarity measure derived from entropy can be generalized by using prior information p_{prior}, different types of probability, e.g. joint probability $p(i_A, i_B)$, conditional probability $p(i_A|i_B)$ or probability of intensity difference $p(i_A - i_B)$.

$$p = (1 - \lambda) \cdot p_{image} + \lambda \cdot p_{prior} \quad ; \quad \lambda \in [0, 1] \quad , \tag{5}$$

where λ is a weighting parameter. Furthermore log function can be replaced with any other function $f(p)$ with strictly monotonically increasing or decreasing first derivative of $p \cdot f(p)$:

$$S^v = f(p); \quad S^r = \frac{1}{N_r} \sum S^v = \frac{1}{N_r} \sum f\,(p). \tag{6}$$

2 Joint distributions

Local multimodal similarity measures are closely related to the joint intensity distribution. To better understand these measures it is necessary to know how the level of misregistration reflects in joint intensity distribution [10].

Suppose we have two simple images A and B representing the same object. 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}). Similarity of two voxels is according to described similarity measures related to the joint probability of intensity pair, so that higher value means better correspondence. When images are correctly registered, the joint distribution contains only two extrema that appear at intensity pairs (i_{1A}, i_{1B}) and (i_{2A}, i_{2B}) as the intensity regions perfectly overlap. When images do not overlap exactly there are some regions with intensity pairs (i_{1A}, i_{2B}) and/or (i_{2A}, i_{1B}), as shown in Fig. 1. The joint probability of mismatched intensity pairs is related to the degree of mismatch where higher mismatch means higher probability. This also means that similarity function based on probability estimation for mismatched regions increases by degree of mismatch. When mismatched regions are larger than correctly matched ones, the estimated similarity of mismatch is higher than similarity for correct match and registration in this case cannot be successful. Due to this fact it is advantageous to use a prior information, from which approximate joint distribution of correctly registered images is known in advance, and therefore the phenomenon mentioned above shall not appear.

In real images tissue types are not represented by a single intensity value, but by a range of intensity pairs, forming an intensity class. Thus, number of intensity classes in one image is equal to the number of biological tissues K. When registering two images, their joint distribution consists of M joint distribution classes C, where each class represents a tissue type pair and all together

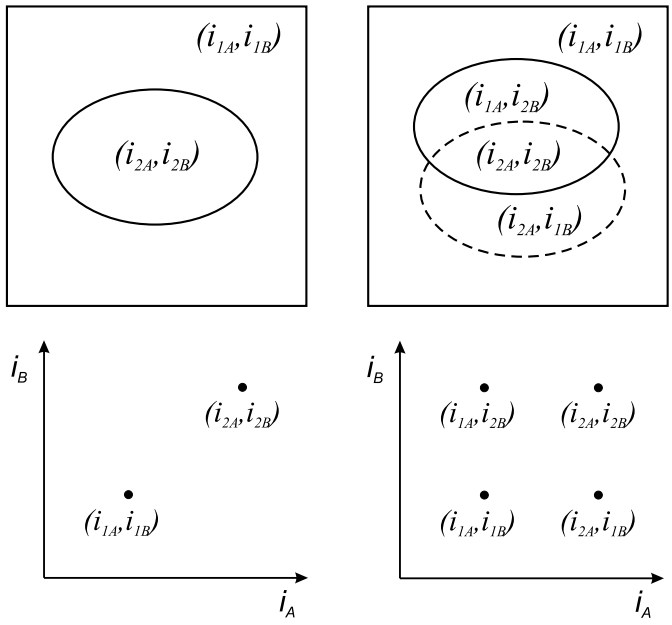

Figure 1. Registered and misregisterd images (top) and their joint distributions (bottom).

form a set $\mathcal{C} = \{C_m\}; m = 1...M$. Among them there are at most K single-tissue classes that correspond to the same tissue type in both images and form a subset $\mathcal{C}_S \subset \mathcal{C}$. All other classes are mixed-tissue classes and correspond to different tissue types in different images. During the registration the amplitude of mixed-tissue classes is expected to decrease until they finally disappear. Joint distribution then consists of only single-tissue classes. If there are no intensity inhomogeneities each class can be modeled by a two dimensional Gaussian function with mean value μ_m, amplitude A_m and covariance matrix $\mathbf{\Sigma}_m$. Class parameters A_m and $\mathbf{\Sigma_m}$ depend on the amount of image noise, where increase of noise reflects to increased class widths and decreased magnitudes. The similarity estimation is getting worse, as the probability of each intensity pair depends only on the number of its occurrences. Actually, image registration does not register biological features, but their intensities. Each tissue type is represented by a whole class of intensity pairs, which are treated independently. If their close relations within the same tissue type were taken into account, they could improve the registration.

Because of insufficient image resolution and image blurring, some voxels represent more than one tissue type. The intensity value i_v of such partial volume voxel v is a weighted sum of class intensities i_k, where weights t_{vk} are portions of tissues $k = 1...K$ that exist at that location. Partial volume voxels are in joint distribution positioned on lines between the classes. Their contribution

to registration is low because of their low probability and thus low estimated similarity. Registration of these voxels is not reliable although edges are most information rich parts of the images. Image processing operations, e.g. linear interpolation and linear image filtering, increase the number of such voxels and therefore worsen the registration. These voxels can be correctly registered only when knowing which classes exists in each voxel and what are their portions.

3 Similarity measures based on segmentation

In order to solve the problems of image noise and partial volume voxels it is necessary to know which intensity pairs form each class and what portion of each partial volume voxel belongs to each class. One possible solution to know this is to perform the segmentation [11]. When registering segmented images similarity measure is perfectly defined by being 1 for the same tissue type on both images and 0 if this is not the case. Considering partial volume voxels, similarity measure of a voxel can be defined as a sum of corresponding voxel portions:

$$S^v = \sum_k \min(t_{kA}, t_{kB}) \quad ; \quad k = 1...K \tag{7}$$

where t_{kA} and t_{kB} are portions of tissue type k that exist in corresponding voxels in image A and B respectively. Unfortunately, image segmentation is not an easy task, either. From the segmentation point of view, images can be segmented by registering to a pre-segmented template. The image registration process is therefore dual to the segmentation. Both of them work better when using the others results. We anticipate that performing both processes simultaneously can improve registration and segmentation results.

We propose a method based on image segmentation to improve registration results. When images are correctly registered there are only single-tissue classes $C_k \in \mathcal{C}_S$, $k = 1...K$, which correspond to image regions with correctly registered tissues. When this is not the case, there are also some mixed-tissue classes that represent different tissues in different images. Without prior information we cannot be absolutely certain which of the classes belong to each group. Therefore, the probabolistic approach can be used. We have found a method to estimate the probability $p(\mathcal{C}_S|C_m)$ that C_m is a single-tissue class. Knowing class parameters μ_m, $\mathbf{\Sigma}_m$ and A_m, for all classes $m = 1...M$, estimated using segmentation methods, the probabilities $p(C_m|\mathbf{i})$ that an intensity pair $\mathbf{i}$ belongs to class C_m can be estimated as well. The sum $\sum p(\mathcal{C}_S|C_m) \cdot p(C_m|\mathbf{i})$ over all classes then represents the probability $p(\mathcal{C}_S|\mathbf{i})$ of $\mathbf{i}$ belonging to one of the single-tissue classes. As each intensity pair in registered images should belong to a single-tissue class, probability $p(\mathcal{C}_S|\mathbf{i})$ can be used as a similarity measure.

3.1 Implementation of segmentation based local similarity measure

Each class C_m; $m = 1...M$ can be modelled by a two-dimensional Gaussian function $p(\mathbf{i}, C_m)$ with mean value μ_m and covariance matrix $\mathbf{\Sigma}_m$. Joint distribution

$p(\mathbf{i})$ can be approximated by a sum

$$p(\mathbf{i}) = \sum_m p(\mathbf{i}, C_m) = \sum_m p(\mathbf{i}|C_m) \cdot p(C_m). \tag{8}$$

It is expected 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 $p(\mathbf{i}, C_l)$; $l = 1...M$, $l \neq m$ can be neglected. Number of classes M, their mean values μ_m and amplitudes A_m can therefore be estimated by extensive search for maxima in joint intensity distribution. When maximum is found, its position is used as a class mean value μ while value itself is used as a class amplitude A. Probability of intensity pairs in $\mathcal{O}_m$ can be approximated by

$$p(\mathbf{i})|_{\mathbf{i}\in\mathcal{O}_m} \approx p(\mathbf{i}, C_m) = A_m \exp\left(-\frac{1}{2}(\mathbf{i}-\mu_m)^T \mathbf{\Sigma}_m^{-1}(\mathbf{i}-\mu_m)\right). \tag{9}$$

By taking a log of (9) we get

$$2\ln\left(\frac{A_m}{p(\mathbf{i})}\right) = (\mathbf{i}-\mu_m)^T \mathbf{\Sigma}_m^{-1}(\mathbf{i}-\mu_m) \quad ; \quad \mathbf{i} \in \mathcal{O}_m, \tag{10}$$

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

$$p(C_m) = \int p(\mathbf{i}, C_m) d\mathbf{i} = 2\pi A_m |\mathbf{\Sigma}_m| \tag{11}$$

The sum of all a priori probabilities C_m, $m = 1...M$, is expected to be 1, but this is not the case even if all class parameters are estimated absolutely correct. The reason is in some intensity pairs with low joint probability that do not belong to any estimated class. Most of such intensity pairs represent partial volume voxels, which are correctly registered only when images are registered with sub-voxel accuracy. To prevent incorrect matching we introduce an additional class C_0 with small constant probability $p(\mathbf{i}, C_0) = \varepsilon$ for each intensity pair. We use ε equal to probability of a single image voxel in joint intensity distribution.

A posterior probability of class C_m, showing the chance of intensity pair to belong to a certain class C_m is according to Bayes rule the following

$$p(C_m|\mathbf{i}) = \frac{p(\mathbf{i}, C_m)}{\sum_{l=0}^{M} p(\mathbf{i}, C_l)}. \tag{12}$$

In order to estimate the similarity as a probability of intensity pair belonging to a single-tissue class $S^v = p(\mathcal{C}_S|\mathbf{i})$, it is necessary to estimate the probability $p(\mathcal{C}_S|C_m)$ that a certain class C_m; $m = 1...M$ is a single-tissue class. Let us assume that each tissue type has unique intensity representation μ. Therefore, among all maxima positioned at the same intensity of image A (or image B)

only one can be a single-tissue class. All the others are mixed-tissue classes, see Fig. 1. Let a subset $\mathcal{C}_{\mu A} \subset \mathcal{C}$ consist of all classes with the same mean value μ_A. Probability $p_A(\mathcal{C}_S|C_m)$ of class C_m being a single-tissue class according to intensity of image A can be estimated as

$$p_A(\mathcal{C}_S|C_m) = \frac{p(C_m)}{\sum\limits_{C_l \in \mathcal{C}_{\mu A}} p(C_l)} \tag{13}$$

This is not a good estimation of probability that class is a single-tissue class, as all classes with higher probability may belong to the same subset $\mathcal{C}_{\mu B}$, which consist of all classes with the same value μ_B. Such situation can be prevented by using probability $p_B(\mathcal{C}_S|C_m)$ that class C_m is a single-tissue class according to intensity of image B, that can be calculated in the same way. The final estimate of probability that class C_m is a single-tissue class is then

$$p(\mathcal{C}_S|C_m) = p_A(\mathcal{C}_S|C_m) \cdot p_B(\mathcal{C}_S|C_m). \tag{14}$$

The sum of $p(\mathcal{C}_S|C_m)$ indicates the number of valid classes. When there are ω single-tissue classes with the same μ_A or μ_B our presumption is incorrect. The estimated probabilities $p_A(\mathcal{C}_S|C_m)$ or $p_B(\mathcal{C}_S|C_m)$ are ω times too small. Anyway, as the relations between the probabilities are still correct registration should still be successful but it may require some more registration steps.

Finally, local similarity based on image segmentation can be estimated as

$$S^v = p(\mathcal{C}_S|\mathbf{i}) = \sum_m p(C_m|\mathbf{i}) \cdot p(\mathcal{C}_S|C_m). \tag{15}$$

4 Comparison results

Comparison of local similarity measures is problematic, as any transformation of a single voxel region makes a huge change in region covering and furthermore the results do not depend only on similarity measure used, but also on spatial model. It appears that the only meaningful way to compare them is their treatment in the context of complete registration systems.

To show the relative performance of our local similarity measures we used two spatially aligned BrainWeb simulated images [12]. The first one, image A (MRI-PD), was used as a reference, while the second one, image B (MRI-T1), was transformed with a known transformation T_0 and then registered back to image A, using multiresolution elastic registration approach, see Fig. 2. The registration result was evaluated using the RMS voxel displacement error e, estimated from difference between voxel displacements achieved by the registration T_R and displacements derived by inverse of known transformation T_0^{-1}. Voxels representing a background were not used for the estimation and were removed by mask Ω, as every registration of the background can be treated as correct. N_Ω denotes number of voxels accepted by the mask Ω and $\mathbf{x}$ are spatial coordinates.

$$e = \sqrt{\frac{1}{N_\Omega} \sum_{\mathbf{x} \in \Omega} \left(T_{0\mathbf{x}}^{-1} - T_{R\mathbf{x}}\right)^2} \quad ; \quad \mathbf{x} = [x_1, x_2, x_3]^T \tag{16}$$

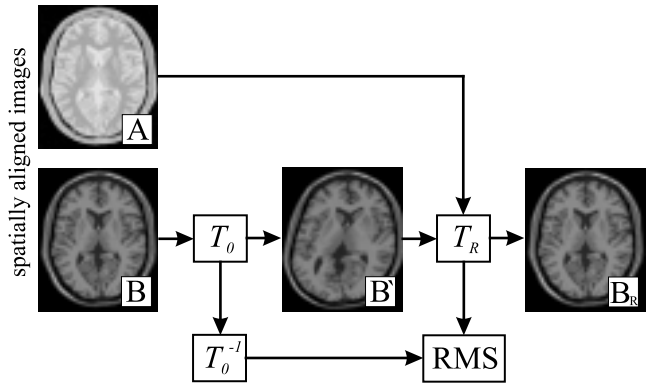

Figure 2. Evaluation scheme.

Table 1. Comparison results of local multimodal similarity measures.

S^v	$e\ [\lambda = 0]$	$e\ [\lambda = 1]$
$p(\mathbf{i})$	6.7449	7.9025
$p(i_A \mid i_B)$	2.6263	2.1834
$p(i_A - i_B)$	6.5745	7.0569
$\log(p(\mathbf{i}))$	2.5631	1.1952
$\log(p(i_A \mid i_B))$	2.8207	1.3988
$\log(p(i_A - i_B))$	3.8948	3.2910
$p(\mathcal{C}_S \mid \mathbf{i})$	**1.2600**	**0.9445**

The results in Table 1 show that using prior information (λ=1) in general improves the registration. Unfortunately, prior information is often not known in advance. The segmentation-based measure is able to distinguish between single-tissue and mixed-tissue classes even without prior information, which makes this result comparable to results of other measures when using prior information. Thus, segmentation-based measure $S^v = p(\mathcal{C}_S|\mathbf{i})$ is shown to be the best among all compared similarity measures. Among local similarity measures derived from entropy original entropy based measure $S^v = \log(p(\mathbf{i}))$ give us the best result while measures based on probability $p(i_A - i_B)$ and measure $S^v = p(\mathbf{i})$ are not suitable for estimation of local properties.

Comparison of estimated similarities derived from the same joint intensity distribution $p(\mathbf{i}) = p(i_A, i_B)$ for entropy based similarity measure $S^v = \log(\mathbf{i})$ and segmentation based similarity measure $S^v = p(\mathcal{C}_S|\mathbf{i})$ are presented in Fig. 3. Entropy based similarity measure differs from joint distribution only in the logarithmic function, which makes maxima wider and their amplitudes more sim-

ilar. Relatively high similarity is achieved even for mixed-tissue intensity pairs, which can lead to misregistration. This is improved in segmentation based measure, where the similarity of mixed-tissue intensity pairs is much lower. The lack of segmentation based measure originates in relatively simple segmentation algorithm that do not detect partial volume voxels, which are then treated as they represent only the most probable tissue type. Such voxels therefore cannot correctly contribute to registration.

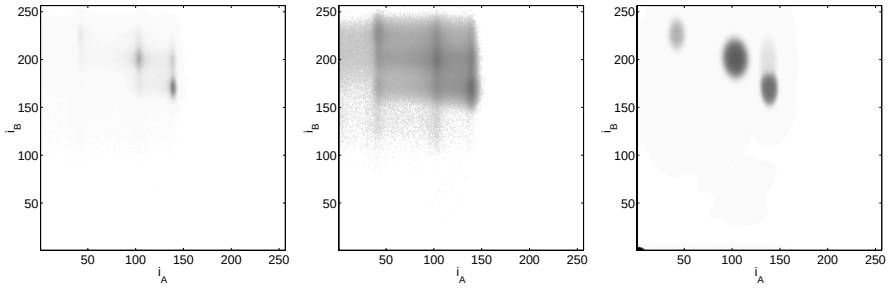

Figure 3. Joint distribution $p(\mathbf{i})$ (left) and estimated similarities: $S^v = \log(p(\mathbf{i}))$ (middle) and $S^v = p(\mathcal{C}_S|\mathbf{i})$ (right).

5 Conclusion

This paper describes some similarity measures for multimodal non-rigid registration. They are all based on probability distribution determined from the whole image content or given in advance as prior information. Measures derived from joint entropy treat each intensity pair individually and do not consider their biological relations. This is improved by segmentation based similarity measures. The method presented in this paper alleviates the problem of noise while problem of partial volume voxels is not directly solved. Partial volume voxels are treated like they belong only to their most probable class. To improve the registration of partial volume voxels, this method needs some improvements.

Segmentation is usually based on presumption that pure single-tissue clusters conform to Gaussian distribution. This is not always the case as real medical images are often influenced by intensity inhomogeneities. Estimated class variances are then to large and registration is not necessarily correct. To alleviate this problem images should be preliminary corrected for intensity inhomogeneities [13]. Such correction is easier to perform when images are already segmented. This indicates that all three tasks, registration, segmentation, and correction for intensity inhomogeneities, should be performed simultaneously.

References

1. 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.
2. 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.
3. J. B. A. Maintz, H. W. Meijering, and M. A Viergever. General multimodal elastic registration based on mutual information. *Medical Imaging*, 3338:144–154, 1998.
4. R. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing*, 46:1–21, April 1989.
5. M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. *Springer Lecture Notes in Computer Science*, 1131:267–276, 1996.
6. 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.
7. 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.
8. I. J. Taneja. On generalized information measures and their applications. *Advances in Electronics and Electron Physics*, 67:327–413, 1989.
9. P. Rogelj and S. Kovacic. Local similarity measures for multimodal image matching. In S. Loncaric, editor, *Proceedings of the first International Workshop on Image and Signal Processing and Analysis – IWISPA 2000*, pages 81–86. University Computing Center, University of Zagreb, 2000.
10. C. Studholme, Hill. D.L.G., and D.J. Hawkes. Using voxel similarity as a measure of medical image registration proc. In E. Hancock, editor, *Proceedings of the British machine vision conference – BMV'94*, pages 235–244, 1994.
11. D.H. Laidlaw, K.W. Fleischer, and A.H. Barr. Partial-volume bayesian classification of material mixtures in mr volume data using voxel histograms. *IEEE Transactions on Medical Imaging*, 17(1):74–86, 1998.
12. 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)*, Lecture Notes in Computer Science, pages 135–140. Springer-Verlag, 1996.
13. B. Likar, J. B. A. Maintz, M.A. Viergever, and F. Pernus. Retrospective shading correction based on entropy minimisation. *Journal of Microscopy*, 197(3):285–295, 2000.