

# Structural Image Representation for Image Registration

Christian Wachinger and Nassir Navab

Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany

wachinge@in.tum.de, navab@in.tum.de

## Abstract

We propose a structural image representation and show its relevance for multi-modal image registration. Structural representation means that only the structures in the image matter and not the intensity values of their depiction. The representation is formulated as a dense descriptor. We specify three properties an optimal descriptor for structural registration has to fulfill: locality preservation, structural equivalence, and discrimination. The proposed entropy images are an approximation to such a representation. We improve their discriminative potential by integrating spatial information in the density estimation. We evaluate entropy images for rigid, deformable, and groupwise multi-modal image registration and achieve very good results in terms of both speed and accuracy. Finally, entropy images seamlessly integrate into existing registration frameworks and allow an efficient registration optimization.

## 1. Introduction

The objective of image registration is to find the correct spatial alignment of images. This task is made difficult by intensity variations between them. Such variations can originate from a multitude of sources, such as illumination changes in optical images, field inhomogeneities in magnetic resonance (MR) images, and, simply, different imaging modalities. A common approach in iconic registration methods is to integrate similarity metrics that are robust to those intensity variations, assuming a functional or statistical intensity relationship, instead of an identical one. On the other hand, geometric registration approaches that build upon an automatic keypoint extraction and description have to apply methods that are robust to intensity variations. Widespread descriptors such as SIFT [8] and GLOH [12] achieve such robustness by building upon intensity differences, rather than absolute intensity values, by calculating histograms of image gradients. The registration of images from different modalities is, however, affected by more substantial intensity variations.

In this article, we present a representation of images that

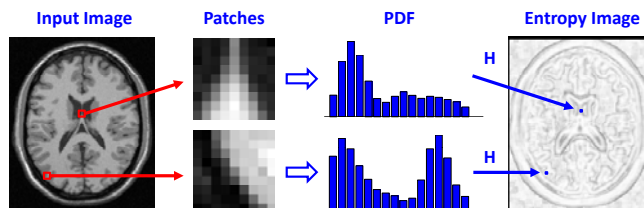


Figure 1. Illustration of the process for calculating entropy images. For each pixel in the image, the local neighborhood patch is selected. For these patches, the PDF of the intensity is estimated, in this example with the Parzen-window method. Finally, the entropy is computed and the result is stored in the corresponding location in order to create the entropy image.

is only dependent on the depicted structures, not on the intensities used to encode them. Such a *structural representation* can assist several image processing tasks. We focus on registration, where this representation enables us to apply fast, mono-modal similarity measures for the alignment of multi-modal images. In our experiments, we measure an average speed-up factor of 6.6 for deformable registration with respect to standard methods. Our process of the creation of such structural images is illustrated in Fig. 1, and can be regarded as a dense descriptor. Examples of structural images are shown in Fig. 2.

We theoretically analyze the properties for a structural representation. We show that the minimal coding length for transferring a patch over a channel, calculated with the *Shannon entropy*, properly captures the information content of a patch invariant to the intensity. However, there is a risk of ambiguities, *i.e.* several patches can lead to the same entropy value. Therefore, we propose to integrate spatial information to the density estimation. Additionally, we thoroughly evaluate the processing steps of the entropy estimation in order to find the best suited representation for registration. The advantages of the structural representation for multi-modal registration are a faster evaluation of the similarity metric and the possibility of using the efficient second-order minimization [2]. In our experiments, we evaluate the application of entropy images for rigid, de-

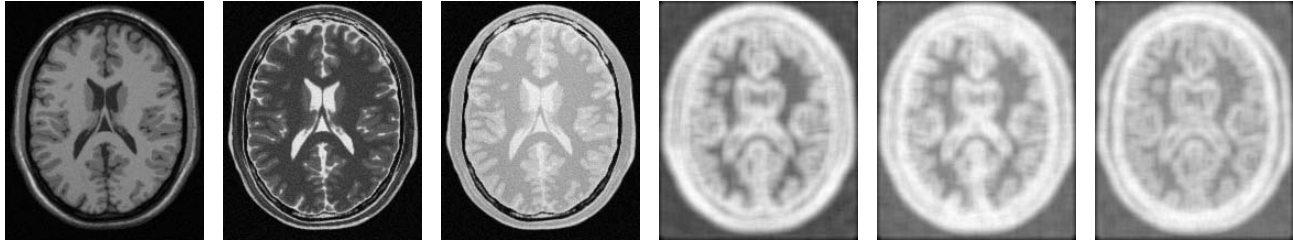


Figure 2. Multi-modal images (T1, T2, PD) from BrainWeb dataset together with entropy images used for rigid registration.

formable, and groupwise registration. We achieve for most multi-modal registration tasks a better accuracy than mutual information, combined with a significant speed-up.

## 2. Related Work

We summarize, first, methods that are related because of the application of entropy for registration, and second, approaches that deal with transforming a multi-modal registration to a mono-modal one. Mutual information (MI) [18, 4] and the congealing framework [7, 19] are widely used techniques for image registration. MI-based methods build upon the entropy calculation of joint and marginal probability distributions, whereas congealing calculates the entropy of a pixel stack of multiple images. Mutual information is applied to gradient images [15] and local phase images [11] for multi-modal registration. In [3], the entropy of the difference image is calculated to align mono-modal images. Also based on the calculation of the entropy is the scale saliency algorithm by Kadir *et al.* [6]. Salient regions are identified in images with the criterion of unpredictability in feature and scale space, where the local entropy is used for quantifying the unpredictability. Our approach is different from [6] because we do not extract keypoint locations but build a dense description.

Techniques that reduce a multi-modal to a mono-modal registration can again be differentiated into two classes. The first ones try to simulate one modality from the other. Examples are X-Ray to CT (Computed Tomography) registration with the creation of digitally reconstructed radiographs [13] and ultrasound to CT registration with the simulation of ultrasound images [17]. The second group consists of methods that transfer both images into a third, artificial modality. Examples are (i) the application of morphological tools [10], (ii) recoloring images depending on the variances of the image regions [1], (iii) the usage of edge- and ridge-information [9], (iv) cross-correlating gradient directions [5], and (v) the creation of shadow-invariant optical images [14]. In our case, we are interested in a general structural representation, so that the application specific approaches are not applicable. Further, (ii) and (iv) use cross-correlation for the comparison, indicating that the descrip-



Figure 3. Patch 1 and 2 show the same structure but encoded with different intensities. All three patches have the same entropy of 2.0749 for an identical spatial weighting. Integrating a location dependent weighting with the modified Gaussian weighting mask shown on the right, we can differentiate patch 3 from the others.

tion is not truly identical. The morphological approach [10] mainly leads to a surface extraction, and although it has gray values instead of only binary values, much internal information is lost. Finally, edge, ridge, and gradient estimation is problematic for points where more than two regions are meeting, *e.g.* T-junctions. This is further discussed in Sec. 5.

## 3. Method

Consider two images  $I, J : \Omega \rightarrow \mathcal{I}$  defined on the image grid  $\Omega$  with intensity values  $\mathcal{I} = \{1, \dots, n\}$ . The registration is formulated as

$$\hat{T} = \arg \max_{T \in \mathcal{T}} \mathcal{S}(I, J(T)), \quad (1)$$

with the space of transformations  $\mathcal{T}$  and the similarity measure  $\mathcal{S}$ . For images with structures being depicted with the same intensity values, so  $I(x) = J(\hat{T}(x))$  for  $x \in \Omega$ , SSD (sum of squared differences) or SAD (sum of absolute differences) are a good choice for  $\mathcal{S}$ . For more complex intensity relationships between the images, such as affine, functional, or statistical ones, typical choices for  $\mathcal{S}$  are the correlation coefficient, correlation ratio, and mutual information, respectively. These are, however, more computationally expensive. Therefore our goal is to find representations of the images so that SSD and SAD become applicable.

Consider patches  $P_x : \mathcal{N}_x \rightarrow \mathcal{I}$  that are defined on the local neighborhood  $\mathcal{N}_x$  around  $x$ . We search a function  $f : P_x \mapsto D(x)$  that assigns each patch a descriptor  $D(x)$  so that the descriptor captures the structural informa-

tion of the patch. The idea is to reduce the problem of finding a structural representation of images to finding such a representation for patches. For example, the patches 1 and 2 in Fig. 3 should lead to the same descriptor. We define two patches  $P_1, P_2$  to be structural equivalent  $P_1 \sim P_2$ , if there exists a bijective function  $g : \mathcal{I} \rightarrow \mathcal{I}$  such that  $\forall y \in \mathcal{N}_x : P_1(y) = g(P_2(y))$ .

Three properties an optimal function  $f$  has to fulfill, with respect to structural registration, are

**P1) Locality preservation:**

$$\|P_i - P_j\| < \varepsilon \implies \|f(P_i) - f(P_j)\| < \varepsilon' \quad (2)$$

**P2) Structural equivalence:**

$$P_1 \sim P_2 \implies f(P_1) = f(P_2) \quad (3)$$

**P3) Discrimination:**

$$P_1 \not\sim P_2 \implies f(P_1) \neq f(P_2) \quad (4)$$

The first property states that similar patches lead to similar descriptors, which is important for the robustness to noise and the capture range. The choice of reasonable  $\varepsilon$  and  $\varepsilon'$  depends on the chosen norm. The second property ensures the structural representation. The third assures a perfect discrimination of patches, which is desirable, however, less critical for the proposed dense descriptor, than it would be for a sparse one.

In order to be able to assess the discrimination ability of the descriptors, we quantify the number of structural different patches. Let  $n = |\mathcal{I}|$  be the number of intensity levels, and  $k = |\mathcal{N}_x|$  be the cardinality of the patch. We assume  $n \geq k$ , which generally holds with typical values for  $n = 256$  and  $k = 10 \times 10$ . For these numbers, we exemplarily indicate the order of magnitude of the patch numbers in the subsequent analysis. The total number of different patches  $\eta_1$  is calculated with

$$\eta_1 = n^k \approx 10^{240}. \quad (5)$$

The number of patches that vary in structure is equivalent to the Bell number  $B$

$$\eta_2 = B(k) = \frac{1}{e} \sum_{l=0}^{\infty} \binom{k}{l} \approx 10^{115}. \quad (6)$$

This corresponds to the number of equivalence classes of the structural equivalence relation  $\sim$ . Patch 1 and 2 in Fig. 3 are in the same class, and are therefore counted only once. The Bell numbers generally indicate the number of ways a set with  $k$  elements can be partitioned into nonempty subsets. This is also the number of patches an optimal function  $f$  would be able to differentiate. However, from a practical point of view, it would require more than 47 bytes per pixel to store, which could exceed the memory limit for volumetric data and decelerate the registration.

### 3.1. Entropy

A function  $f$  that is building upon calculating the mean value or differences between the entries in the patch (gradient), is not suitable for extracting the structural information. A more general concept is to quantify the information content of a patch or, analogously, the bound for a lossless compression, as stated by Shannon's theorem, which is both achieved with the entropy. The Shannon entropy of a random variable  $Y$  with possible values  $\mathcal{I}$  is

$$H(Y) = - \sum_{i \in \mathcal{I}} p(Y = i) \cdot \log p(Y = i), \quad (7)$$

assuming  $p$  to be the probability density function (PDF) of  $Y$ . Calculating the entropy on a dense image grid leads to

$$D_I(x) = H(I(\mathcal{N}_x)). \quad (8)$$

The construction of *entropy images* is illustrated in Fig. 1.

We analyze if the entropy fulfills the listed properties P1 to P3. Consider two intensity distributions  $p_1$  and  $p_2$ , corresponding to patches  $P_1$  and  $P_2$ , respectively. Small changes in the patches also lead to small changes in the distributions, so that the locality preservation is fulfilled. The second property, however, does not hold. The number of different distributions is

$$\eta_3 = \binom{n+k-1}{k} \approx 10^{90}, \quad (9)$$

which corresponds to ball picking of unordered samples with replacement.

In order to make distributions fulfill also the second property, we have to sort the entries of the distribution

$$p' = \text{sort}(p). \quad (10)$$

The number of sorted distributions  $p'$  is

$$\eta_4 = \mathcal{P}(k) \approx \frac{1}{4k\sqrt{3}} e^{\pi\sqrt{2k/3}} \approx 10^8 \quad (11)$$

with the partition function  $\mathcal{P}$ , which represents the number of ways of writing an integer as a sum of positive integers, where the order of addends is not considered significant.

The final step, the mapping from ordered histograms to real values, is performed with the entropy formula in Eq. (7). For  $k = 2$ , the entropy uniquely assigns each ordered histogram a scalar. However, for  $k \geq 3$  the mapping is no longer injective and consequently  $\eta_4$  presents an upper bound to the number of different entropy values for patches. Generally, the desired number  $\eta_2$  is much higher than the maximally achievable  $\eta_4$ . To address this issue, we propose to add a spatial weighting in the density estimation in Sec. 4.2, leading to an improved discrimination.

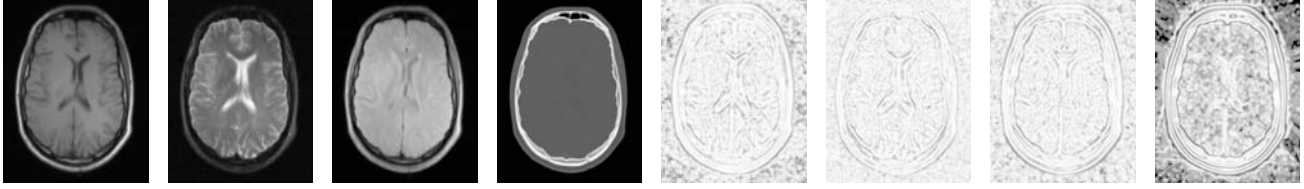


Figure 4. Multi-modal images (T1, T2, PD, CT) from RIRE dataset together with entropy images used for deformable registration.

### 3.2. Multi-Modal Registration with Entropy Images

In contrast to the regular registration in Equ. (1), the formulation of registration based on entropy images is

$$\hat{T}' = \arg \max_{T \in \mathcal{T}} \mathcal{S}(D_I, D_J(T)), \quad (12)$$

with  $\mathcal{S}$  being SSD or SAD, even for images from different modalities. Both registration problems are obviously related, because the entropy images are calculated on the original images, but the results of the registrations,  $\hat{T}$  and  $\hat{T}'$ , do not necessarily have to be identical.

We would like to point out that the structural representation is different to a *modality invariant* one. Each imaging device has its own characteristics, leading to images with specific artifacts and noise. Also, structures visible in one of the images may not be observable in the second one. As an example, compare the CT and T2 image in Fig. 4, whose appearance is significantly different. It is also clear that entropy images cannot detect structures, where there are none. However, we can expect to be robust to such changes, so that those structures that are present in all images can guide the registration. This problem is not specific to entropy images, but rooted in the multi-modal registration scenario, and therefore also affects multi-modal metrics like MI.

### 3.3. Efficient Pair- and Groupwise Registration

One part for a fast registration is the discussed similarity measure, but not less important is an efficient optimization method. The efficient second-order minimization (ESM) [2] is an extension of Gauss-Newton (GN) and was shown to converge faster than other gradient-based optimizers [2, 16]. ESM builds a second-order approximation of the cost function without the explicit calculation of the second derivative. This is achieved by combining the image gradients of the moving and fixed image  $\nabla I + \nabla J$ . We refer to the mentioned references for further details. For multi-modal images this combination is not meaningful, and consequently, ESM has so far not yet been applied for registering them. The structural representation of images, however, enables us to apply ESM for multi-modal images.

In [16], we deduced ESM for groupwise registration. For the case of symmetric similarity measures,  $\frac{N \cdot (N-1)}{2}$  pairwise measures have to be calculated for simultaneous registration, with  $N$  the number of images. Consequently, the

influence of the speed-up of simple matching functions is quadratic. Concluding the effects of entropy images for multi-modal registration, they enable the usage of an efficient optimizer, needing less steps, and further, they permit a faster calculation of each update step.

## 4. Details on Entropy Estimation

There are several processing steps in the entropy estimation that influence the appearance of the entropy image, compare Figs. 2 and 4. We evaluate their effects for rigid and deformable registration. Both types of registration have different requirements and, consequently, lead to different preferences. We present details about the experimental setup for the evaluation in Sec. 5.

### 4.1. Local Neighborhood

For rigid registration, we evaluate local neighborhoods ranging from  $5 \times 5$  to  $19 \times 19$  pixels. Patch sizes of  $5 \times 5$  and  $7 \times 7$  lead to unsatisfying results. When further increasing the patch size, the similarity plots become wider, because the entropy images are smoother. For our rigid experiments, we continue with patch sizes of  $11 \times 11$ , as a good compromise between computational complexity and sufficient statistics for an accurate estimation. For deformable registration, we prefer smaller patch sizes to have a more local description, and therefore test them in the range from  $5 \times 5$  to  $13 \times 13$ , with good results for  $7 \times 7$  patches. In 3D, we achieved good results with  $9 \times 9 \times 9$  patches, where we adapt the neighborhood for data with anisotropic spacing.

### 4.2. Spatially-weighted PDF Estimation

We use histogramming and the kernel-based Parzen window method for the non-parametric PDF estimation, with the latter yielding more robust results for a small number of samples. For both, the bin size has to be specified. A large number of bins makes the entropy image more sensitive to noise, while a low number deteriorates the unique patch description.

As discussed previously, we are interested in increasing the discrimination of the entropy estimation, because of the large difference between  $\eta_2$  and  $\eta_4$ . Consider, for instance, the three patches in Fig. 3. While it is desired to assign patch

1 and 2 the same value, this does not hold for patch 3. However, the PDF is the same under permutation of pixels in the patch. To address this issue, we propose to modify the density estimation with plug-in estimators. When constructing the image histogram  $h$ , each pixel's intensity is added with a constant summand to the histogram. We introduce a spatial weighting function  $\omega : \mathcal{N}_x \rightarrow \mathbb{R}$ , assigning a weight to each patch location. The histogram update changes to

$$\forall y \in \mathcal{N}_x : h_x[I(y)] \leftarrow h_x[I(y)] + \omega(y). \quad (13)$$

The Parzen window density estimation changes accordingly. The PDF is obtained by convolution of the histogram with a Gaussian kernel  $G_\sigma$  of standard deviation  $\sigma$

$$p_x = \frac{1}{M} (G_\sigma * h_x) \quad (14)$$

and normalization with  $M = \sum G_\sigma * h_x$ . In our experiments, we use a Gaussian, a modified Gaussian, and the identity as weighting functions. The identity corresponds to the usual density estimation. For the Gaussian we set  $\omega(y) = G_\sigma(\|y - c\|)$  with  $c$  the patch center. The discrimination between patches is not optimal because the Gaussian is rotational symmetric. We therefore modify the Gaussian weighting function, see Fig. 3, giving it a unique weight at each patch location. We assign similar values to neighboring locations to ensure the locality preservation.

Although each location has a different weight for the modified Gaussian, the sum of several values, as it is done in the histogram calculation, can lead to the same value, and therefore ambiguities. An optimal assignment of weights to the  $|\mathcal{N}_x|$  patch locations in the weighting mask, so that they are unique with respect to addition and that the dynamic range is minimal, is to use  $2^{-1}, 2^{-2}, \dots, 2^{-|\mathcal{N}_x|}$ . However, even then the dynamic range is too high, considering  $2^{-49} \approx 10^{-15}$ , leading to locations that become negligible in the entropy calculation.

### 4.3. Intensity Normalization

In order to use the whole range of the histogram, we normalize the intensity values of patches  $P_x$ . For this, we can either use the global minimum and maximum

$$\min_x = \inf_{y \in \Omega} I(y) \quad \max_x = \sup_{y \in \Omega} I(y) \quad (15)$$

or the local extrema in the patch

$$\min_x = \inf_{y \in \mathcal{N}_x} I(y) \quad \max_x = \sup_{y \in \mathcal{N}_x} I(y). \quad (16)$$

Our experiments confirm that a global approach is better suited for rigid, while the local approach is better suited for deformable registration.

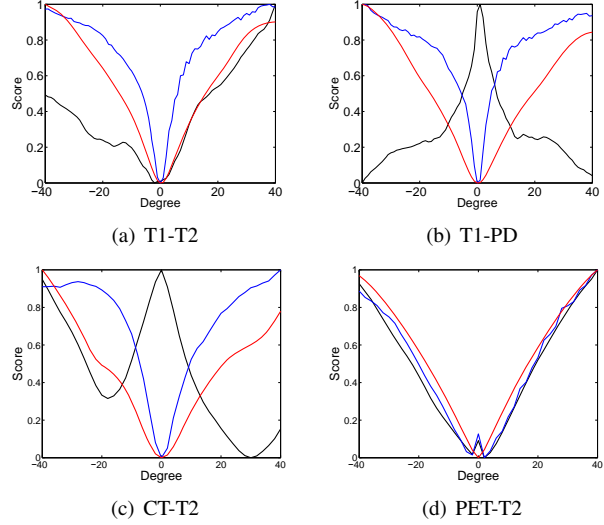


Figure 5. Similarity plots for RIRE dataset (black: SSD, blue: MI, red: eSSD).

## 5. Experiments

We conduct experiments on T1-, T2, and PD-weighted MR images from the BrainWeb database<sup>1</sup> and CT, T1, T2, PD, and PET (Positron Emission Tomography) images from the Retrospective Image Registration Evaluation (RIRE) database<sup>2</sup>. For both datasets the ground truth alignment is provided. We depict axial slices of the original and entropy images in Figs. 2 and 4. The average time for the creation of an entropy image in C++ is 0.078s for a slice of  $256 \times 256$ , and 11.01s for a volume of  $512 \times 512 \times 29$ .

For our deformable experiments we use the freely available deformable registration software DROP<sup>3</sup>, which applies free-form deformations based on cubic B-Splines and an efficient linear programming with a primal-dual scheme. To emphasize that we consider deformable registration as an available tool, we perform all experiments with the standard settings of DROP, without any optimization towards the application of entropy images. This should also guarantee that similar results are obtainable with alternative registration approaches.

### 5.1. Rigid Registration

For rigid registration, the standard configuration for the entropy images is:  $11 \times 11$  patches (in 2D) and  $9 \times 9 \times 9$  patches (in 3D), 64 bins, Gaussian weighting, global normalization, Parzen-window estimation, and Shannon entropy. We evaluate the usage of SSD, MI, and SSD on entropy images (eSSD) for the rigid alignment by analyzing the similarity plots for the various multi-modal combina-

<sup>1</sup><http://www.bic.mni.mcgill.ca/brainweb/>

<sup>2</sup><http://www.insight-journal.org/rire/>

<sup>3</sup><http://www.mrf-registration.net>

Table 1. Pairwise registration Study for RIRE. Errors for translation in mm and rotation in degree.

Data Set	Sim	$t_x$	$t_y$	$t_z$	$r_x$	$r_y$	$r_z$	$t_{\text{total}}$	$r_{\text{total}}$
T1-T2	MI	0.719	0.395	1.531	1.594	2.252	1.467	1.754	3.139
T1-T2	eSSD	0.042	0.224	0.396	1.120	1.582	0.538	<b>0.461</b>	<b>2.013</b>
T1-PD	MI	0.190	0.251	0.856	0.635	0.877	0.546	0.944	1.231
T1-PD	eSSD	0.061	0.048	0.431	0.344	0.648	0.558	<b>0.442</b>	<b>0.926</b>
CT-T1	MI	1.925	1.004	1.312	1.718	2.951	0.763	2.710	3.951
CT-T1	eSSD	0.963	1.269	0.702	2.433	0.728	0.169	<b>1.997</b>	<b>3.089</b>
PET-T1	MI	9.071	7.730	13.409	29.226	23.578	4.945	20.869	46.234
PET-T1	eSSD	0.053	0.057	0.089	0.040	0.038	0.042	<b>0.135</b>	<b>0.078</b>

tions of both datasets, see Fig. 5. The plots are created by rotating the images around the image center for the range of  $[-40^\circ, 40^\circ]$ , with  $0^\circ$  corresponding to the correct alignment. Generally, we observe that SSD fails, which was to be expected, whereas MI and eSSD indicate the correct alignment. In most cases eSSD provides smoother curves with a wider basin than MI, which is advantageous for registration, because it enables a larger capture range. Remarkable is also the problem of MI for the registration of PET images.

We further run a pairwise registration study for the various combinations of the multi-modal volumes, with an initial random deviation of maximal  $\pm 20$  mm in translation and  $\pm 20^\circ$  in rotation from the correct pose, to compare the performance of eSSD and MI. For each configuration 100 registrations are performed using the best neighbor optimizer. In Tab. 1, we show the absolute mean error for each pose parameter and the total mean error for translation and rotation for the RIRE dataset. On the BrainWeb dataset eSSD and MI lead to comparable results, where on the MR volumes of the RIRE dataset, eSSD performs significantly better, matching with our observations from the similarity plots. For the alignment with CT volumes the registration error increases for both eSSD and MI. The experiments for the PET registration are performed on volumes with a lower resolution to match the resolution of the PET volume ( $128 \times 128 \times 29$ ). The registration with eSSD achieves excellent results, in contrast to MI.

We measure an average computation time for SSD being a factor of 15 faster than for MI. Moreover, since SSD is better suited for parallelization than MI, a GPU implementation could lead to further benefits.

## 5.2. Deformable Registration

We first evaluate the application of gradient direction (GD) [5] and gradient magnitude (GM) for deformable registration. GM achieved the best results in a comparison of edge- and ridge-based methods [9]. GD and GM have problems with several areas meeting in one point, such as T-junctions, as it is illustrated for two synthetic images in

Fig. 6. We overlay the gradient fields from both images, in red and blue, on the second image. It becomes clear that the directions are completely different at the junction, so that also the in [5] proposed normalization cannot correct this. Further, the representation with GM changes with the intensity values. Contrary, the entropy images correctly represent the structure of the images, leading to good registration results.

On the medical databases, we deform one of the two images with a deformation  $d_g$  serving as ground truth. Next, we run the registration with the deformed image as target and the image from the other modality as source to calculate the estimated deformation  $d_c$ . We calculate the average euclidean difference of the deformation fields  $\tau = \frac{1}{|\Omega|} \sum_{x \in \Omega} \|d_c(x) - d_g(x)\|$  for quantifying the residual error of the registration.

The results for the experiments are shown in Fig. 7 and Tab. 2. The error of the registration with the original images using normalized mutual information is stated in the table as reference. The standard configuration for the entropy image for deformable registration is:  $7 \times 7$  patches, 16 bins, Gaussian weighting, local normalization, Parzen-window estimation, and Shannon entropy. In our experiments, we evaluate each of the parameters by changing one of them from the standard configuration and letting the others constant.

From Fig. 7, we see that best results are achieved around 16 bins. While reducing it further to 8 bins also leads to good results, increasing it further to 32 bins leads to an increase in error. A good compromise in the patch size for the different datasets and modalities is approximately  $7 \times 7$ . Larger patch sizes still lead to good results on the RIRE images, but on the Brainweb images we observe a significant increase of the error. Smaller patches lead to an inaccurate density estimation because of the small number of samples. For the weighting, we observe a general reduction of the error when using a more advanced weighting than the standard identical one.

From Tab. 2, we see that a local normalization of the

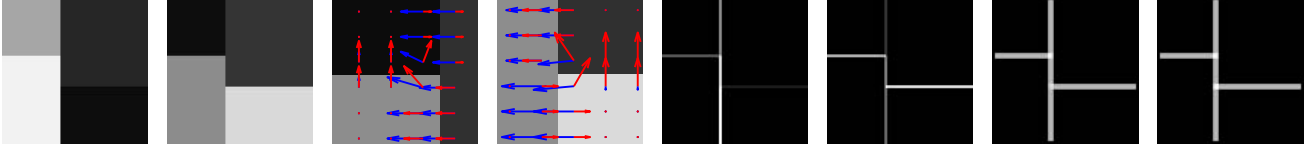


Figure 6. Illustration of two synthetic multi-modal images ( $1^{st}$  &  $2^{nd}$ ) together with a zoom on gradient fields ( $3^{rd}$  &  $4^{th}$ ), gradient magnitude images ( $5^{th}$  &  $6^{th}$ ) and entropy images ( $7^{th}$  &  $8^{th}$ ). The best structural representation is achieved with entropy images.

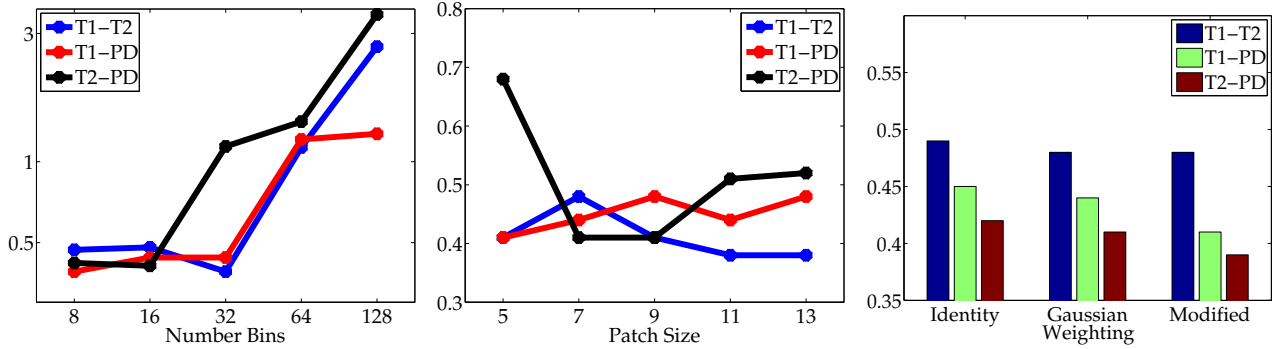


Figure 7. Evaluation of error  $\tau$  for deformable registration for varying bin number, patch size, and spatial weighting on RIRE.

intensity values and the Parzen window method for the density estimation are essential for good deformable registration results. For the entropy estimation, we obtain best results for the Shannon entropy. The results for the special Rényi entropy  $H_2$  are comparable, while the ones for the Burg entropy are not good.

The results of the deformable registration on T1, T2, and PD images show a slight advantage for normalized mutual information (NMI) on the Brainweb dataset, while for the RIRE dataset the entropy images lead to a significant improvement. The registration with CT and PET is more challenging, because of the significant differences in the images. The registration of CT with entropy images is inferior to NMI. For the registration of PET, entropy images are superior to NMI.

The registration with eSSD is in average 6.6 times faster than with NMI. This includes the time for the creation of the entropy images.

### 5.3. Groupwise Registration

Finally, we perform a simultaneous, rigid registration study for the three BrainWeb and RIRE MR volumes. We compare the usage of ESM with eSSD, Gauss-Newton with eSSD, and Gauss-Newton on the original images with MI. We run 50 registrations, each starting from a random initial position. Each initial position has an accumulated RMS error of 45 over all volumes from the correct alignment, weighting 1mm equal to  $1^\circ$ . The average residual error for each step is shown in Fig. 8. We observe that ESM converges significantly faster than GN. For BrainWeb, GN on

the original and entropy images are comparable. For RIRE, GN on the entropy images works well. For GN on the original dataset with MI, however, most registrations do not converge, confirming our previous experiments.

## 6. Conclusion

We proposed entropy images as a structural representation for images, and validated their relevance for image registration. Entropy images are not optimal, because the discrimination is below the theoretical maximum, however, we also showed that such an optimal representation, if known, would yield practical problems. We further analyzed each of the processing steps for the entropy calculation with the objective to find the best components. Among others, we proposed the integration of a spatial weighting into the density estimation. We evaluated the performance of entropy images on multiple datasets for rigid, deformable, and groupwise registration. The alignment accuracy is in almost all cases comparable or better than for mutual information, with a significant registration speed-up. Finally, entropy images allow us to use the efficient ESM optimizer for multi-modal registration, where our experiments confirm its fast convergence rate.

## 7. Acknowledgment

We thank Ramtin Shams for helping with the experiments. This work was partly funded by the European Project "PASSPORT".

Table 2. Registration errors  $\tau$  in mm for various configurations for calculating the entropy images. (B: Brainweb, R: RIRE dataset)

Technique	T1-T2 <sub>B</sub>	T1-PD <sub>B</sub>	T2-PD <sub>B</sub>	T1-T2 <sub>R</sub>	T1-PD <sub>R</sub>	T2-PD <sub>R</sub>	CT-T1 <sub>R</sub>	PET-T2 <sub>R</sub>
NMI (reference)	0.63	0.79	0.66	0.94	1.04	1.33	1.84	3.42
Local Norm.	0.42	0.58	0.56	0.48	0.44	0.41	5.86	0.64
Global Norm.	0.99	2.08	0.87	2.76	4.30	4.11	6.87	1.83
Parzen Window	0.42	0.58	0.56	0.48	0.44	0.41	5.86	0.64
Histogramming	0.54	0.91	0.66	0.94	1.12	1.42	6.02	0.97
Shannon	0.42	0.58	0.56	0.48	0.44	0.41	5.86	0.64
Rényi, $H_2$	0.47	1.11	0.64	0.49	0.54	0.64	6.39	1.25
Burg	1.82	4.61	2.43	2.81	2.37	2.68	6.71	3.46

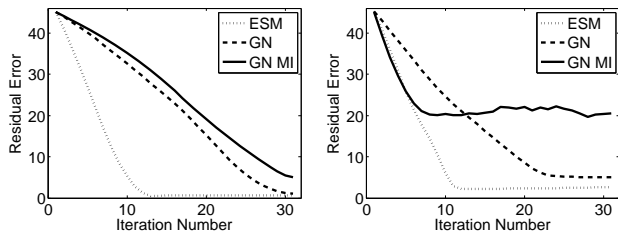


Figure 8. Groupwise convergence study for BrainWeb (left) and RIRE (right) volumes.

## References

- [1] A. Andronache, M. von Siebenthal, G. Székely, and P. Cattin. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Medical Image Analysis*, 12(1):3–15, 2008.
- [2] S. Benhimane and E. Malis. Real-time image-based tracking of planes using efficient second-order minimization. In *IEEE/RSJ*, pages 943–948, 2004.
- [3] T. Buzug, J. Weese, C. Fassnacht, and C. Lorenz. Image registration: Convex weighting functions for histogram-based similarity measures. *Lecture Notes in Computer Science*, pages 203–212, 1997.
- [4] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3D multi-modality medical image registration using feature space clustering. In *Computer Vision, Virtual Reality and Robotics in Medicine*, 1995.
- [5] E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. *Methods of information in medicine*, 46(3):292–299, 2007.
- [6] T. Kadir and M. Brady. Saliency, scale and image description. *Int. J. Comput. Vision*, 45(2):83–105, 2001.
- [7] E. G. Learned-Miller. Data driven image models through continuous joint alignment. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 28(2):236–250, 2006.
- [8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.
- [9] J. Maintz, P. van den Elsen, and M. Viergever. Comparison of edge-based and ridge-based registration of CT and MR brain images. *Medical image analysis*, 1(2):151–161, 1996.
- [10] J. Maintz, P. van der Elsen, and M. Viergever. Registration of 3D Medical Images Using Simple Morphological Tools. In *International Conference on Information Processing in Medical Imaging*, pages 204–217, 1997.
- [11] M. Mellor and M. Brady. Phase mutual information as a similarity measure for registration. *Medical Image Analysis*, 9(4):330–343, 2005.
- [12] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(10):1615–1630, 2005.
- [13] G. Penney, J. Weese, J. Little, P. Desmedt, D. Hill, and D. Hawkes. A comparison of similarity measures for use in 2-d-3-d medical image registration. *Medical Imaging, IEEE Transactions on*, 17(4):586–595, Aug. 1998.
- [14] D. Pizarro and A. Bartoli. Shadow resistant direct image registration. In *SCIA*, pages 928–937, 2007.
- [15] R. Shams, R. A. Kennedy, P. Sadeghi, and R. Hartley. Gradient intensity-based registration of multi-modal images of the brain. In *ICCV*, Oct. 2007.
- [16] C. Wachinger and N. Navab. Similarity Metrics and Efficient Optimization for Simultaneous Registration. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.
- [17] W. Wein, S. Brunke, A. Khamene, M. Callstrom, and N. Navab. Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. *Medical Image Analysis*, 12:577–585, Oct. 2008.
- [18] W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1996.
- [19] L. Zöllei, E. Learned-Miller, E. Grimson, and W. Wells. Efficient population registration of 3d data. In *Computer Vision for Biomedical Image Applications, ICCV*, 2005.