Entropy and Laplacian images: Structural representations for multi-modal registration

Christian Wachinger *, Nassir Navab

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

ARTICLE INFO

ABSTRACT

The standard approach to multi-modal registration is to apply sophisticated similarity metrics such as mutual information. The disadvantage of these metrics, in comparison to measuring the intensity difference with, e.g. L1 or L2 distance, is the increase in computational complexity and consequently the increase in runtime of the registration. An alternative approach, which has not yet gained much attention in the literature, is to find image representations, so called structural representations, that allow for the application of the L1 and L2 distance for multi-modal images. This has not only the advantage of a faster similarity calculation but enables also the application of more sophisticated optimization strategies. In this article, we theoretically analyze the requirements for structural representations. Further, we introduce two approaches to create such representations, which are based on the calculation of patch entropy and manifold learning, respectively. While the application of entropy has practical advantages in terms of computational complexity, the usage of manifold learning has theoretical advantages, by presenting an optimal approximation to one of the theoretical requirements. We perform experiments on multiple datasets for rigid, deformable, and groupwise registration with good results with respect to both, runtime and quality of alignment.

1. Introduction

The objective of image registration is to find the correct spatial alignment between corresponding structures in images. This task is made difficult by intensity variations between images. 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 (Lowe, 2004) and GLOH (Mikolajczyk and Schmid, 2005) 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 introduce a representation of images that is only dependent on the depicted structures and not on the intensities used to encode them. Such a structural representation can assist several image processing tasks, while we focus on registration in this article. We obtain structural representations by calculating a dense set of descriptors that capture the structural information of each of the local patches. Subsequently, the input images are replaced by the dense set of descriptors, on which a regular intensity-based registration is performed. This guarantees a seamless integration into existing registration frameworks. The advantage of the structural representation is that the simple L1 or L2 distance can be used for the registration of multi-modal images. These metrics are computationally less expensive than mutual information and allow therefore a faster registration. This is even more important for groupwise registration, where the speed improvement becomes quadratic, as we discuss in Section 3.2. Moreover, an efficient optimization scheme, the efficient second-order minimization (ESM, Benhimane and Malis, 2004; Wachinger and Navab, 2009), can be applied for multi-modal registration. Finally, L1 and L2 distances are better suited for parallelization than complex multi-modal similarity measures, which is important with respect to transferring the computation to GPUs for further speed-up. A preliminary of version of this work was presented at conferences (Wachinger and Navab, 2010b; Wachinger and Navab, 2010a). This article expands the conference papers with a more detailed

* Corresponding author.
  E-mail addresses: wachinge@in.tum.de (C. Wachinger), navab@in.tum.de (N. Navab).

description and differentiation of both approaches, together with more extensive experiments.

In Sections 3 and 4, we describe the integration of structural images for registration and theoretically analyze the properties of a structural representation. In Section 5, 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. The created entropy images therefore present an exemplary structural representation. There is, however, a risk of ambiguities, *i.e* several patches can lead to the same entropy value. In order to address this issue, we propose to integrate spatial information to the density estimation. In Section 6, we introduce an alternative structural representation, based on the application of manifold learning. The representation created with Laplacian eigenmaps has superior theoretical properties, because it optimally fulfills one of the requirements for a structural representation, the preservation of locality. This means that patches, which are close in high-dimensional patch space, are mapped to a close structural representation. Moreover, this technique exploits the internal similarities across modalities, similar to Penney et al. (2008), for the structural embedding. We refer to the images created with Laplacian eigenmaps shortly as *Laplacian images*. In Sections 7 and 8, we evaluate the performance of entropy and Laplacian images.

2. Related work

There are two groups of related work, first, methods that are related because they deal with transforming a multi-modal registration to a mono-modal one, and second, articles that are related to our proposed representations (see Sections 2.1 and 2.2).

Techniques that reduce a multi-modal to a mono-modal registration can 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 (Penney et al., 1998) and ultrasound to CT registration with the simulation of ultrasound images (Wein et al., 2008). In our case, we are interested in a general structural representation, so that the application of these specific approaches is not applicable.

The second group consists of methods that transfer both images into a third, artificial modality. Examples are (i) the application of morphological tools (Maintz et al., 1997), (ii) recoloring images depending on the variances of the image regions (Andronache et al., 2008), (iii) the usage of edge- and ridge-information (Maintz et al., 1996), (iv) cross-correlating gradient directions (Haber and Modersitzki, 2007), and (v) the creation of shadow-invariant optical images (Pizarro and Bartoli, 2007). Approaches (ii) and (iv) use cross-correlation for the comparison, indicating that the description is not truly identical. The morphological approach (Maintz et al., 1997) mainly leads to a surface extraction, and although it employs 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, as discussed in Section 7.

2.1. Entropy images

Concepts from information theory, specifically the entropy, have a significant influence on image registration. The widely utilized mutual information (MI) (Wells et al., 1996; Collignon et al., 1995) is building upon the entropy calculation of joint and marginal probability distributions. Similarly, the congealing framework (Learned-Miller, 2006; Zöllei et al., 2005), which is commonly used for the simultaneous alignment of multiple images, evaluates the entropy

of a pixel stack. In Buzug et al. (1997), 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 (Kadir and Brady, 2001). 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.

2.2. Laplacian images

As we will describe in more details in Section 6, the application of Laplacian images for multi-modal registration relies on the assumption of similar internal similarities in images across modalities. This was previously exploited in a novel framework for multi-modal registration based on internal similarity (Penney et al., 2008). The two approaches are similar because small image patches are compared to find local similarities inside images. The consecutive utilization of this information is, however, entirely different. While in Penney et al. (2008) a few internal similarity structures are identified and then transferred to the image from the other modality, we use all the patch information to build a neighborhood graph, approximating the patch manifold embedded in high dimensions. The internal similarity in images is also exploited in Shechtman and Irani (2007), however, only locally for constructing image descriptors.

Finally related is the approach for learning similarity measures for multi-modal registration (Lee et al., 2009). For the supervised learning a max-margin structured output learning is used. The approach seems related because also learning is considered, however, we are interested in finding structural representations, instead of learning the similarity measure, and our unsupervised learning with Laplacian eigenmaps does not need any training. The training is not simple because correctly aligned images from the target modalities must be available. It is even more challenging for magnetic resonance (MR) images, which vary significantly in their appearance for different echo and repetition times (TE/TR).

3. Structural image registration

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

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

with the space of transformations $\mathcal{T}$ and the similarity measure $\mathcal{S}$. For a more complex intensity relationship than the identity, such as an affine, a functional, or a statistical one, typical choices for $\mathcal{S}$ are the correlation coefficient, correlation ratio (Roche et al., 1998), and mutual information, respectively. These are, however, more computationally expensive than the L2 distance. Our goal is therefore to find structural representations D^I and D^J that replace I and J in the optimization of Eq. (1) and for which we can apply mono-modal metrics. The registration is then formulated as follows

$$\widehat{T} = \arg\max_{T \in \mathcal{T}} \ \mathcal{S}(D^I, D^J(T)), \tag{2}$$

with $\mathcal{S}$ being the L1 or L2 distance, even for images from different modalities. We illustrate the entire process schematically in Fig. 1.

3.1. Efficient optimization

One part of a fast registration is the discussed similarity measure, but not less important is an efficient optimization scheme. The efficient second-order minimization is an extension of Gauß–Newton (GN) and was shown to converge faster than other gradient-based optimizers (Benhimane and Malis, 2004;
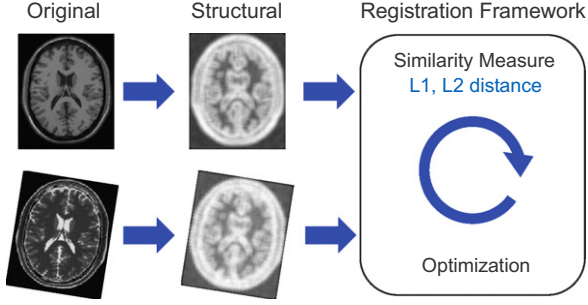
Fig. 1. Schematic illustration of structural registration. From the original images, structural representations are calculated. In this diagram, entropy images are shown. Subsequently, these images are used in the standard intensity-based registration framework, with L1 or L2 distance as possible similarity measures.

Wachinger and Navab, 2009). The efficient second order minimization (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 of image gradients is not meaningful, and consequently, ESM has not yet been applied to multi-modal registration. The structural representation of images, however, enables us to apply ESM for multi-modal images, as shown in Section 7. Further for deformable registration, the application of MI-based similarity measures causes non-sparse matrices for GN, leading to high computational complexity (Modersitzki, 2009, Chapter 9).

There are also other optimization schemes like the registration algorithms based on the Fourier transform (Reddy and Chatterji, 1996; Padfield, 2010) or the batch alignment by sparse and low-rank decomposition (Peng et al., 2010) that can benefit from structural images. Fourier-based approaches calculate the correlation between images and assume linearly correlated images, analogously to Peng et al. (2010). This assumption is generally too limiting for multi-modal registration. The application of structural representations has the potential to make those approaches even applicable to multi-modal registration.

3.2. Efficient groupwise registration

For groupwise registration it is even more important to have an efficient registration process, since the computational cost increases significantly when moving to the alignment of multiple images. There exist various techniques, such as the congealing framework (Learned-Miller, 2006) or the accumulation of pairwise estimates (APE, Wachinger and Navab, 2009; Wachinger et al., 2007), to measure the similarity of a group of images. In Wachinger and Navab (2009), the derivation of ESM for APE is presented. Assuming a group of N images $\{I_1, \ldots, I_N\}$, APE is formulated as

$$\text{APE}_{\mathcal{S}}(I_1, \ldots, I_N) = \sum_{i=1}^{N} \sum_{\substack{j=1, j \neq i}}^{N} \mathcal{S}(I_i, I_j). \tag{3}$$

In this case, $N \cdot (N-1)$ pairwise similarity measures have to be calculated for simultaneous registration. Consequently, the influence of the speed-up of simple matching functions is quadratic. To conclude, structural representations for multi-modal registration enable the usage of an efficient optimizer, needing less steps, and further, they permit a faster calculation of each update step.

4. Structural representation

In this section, we analyze the theoretical requirements on a structural representation. We break the problem of finding a struc-

tural representation for images down to the simpler problem of finding a structural representation for image patches. We denote patches that are defined on the local neighborhood $\mathcal{N}_x$ around x as $P_x : \mathcal{N}_x \to \mathcal{I}$. Our objective is to find a function $f : P_x \mapsto D_x$ that assigns each patch a descriptor D_x so that the descriptor captures the structural information of the patch. Since we calculate a descriptor for each location x, we obtain a new image with the original intensities replaced by the descriptors. Moreover, we differentiate between patches of different images, with P_x being part of I and Q_x being part of J, and further $f' : Q_x \mapsto D'_x$.

We define two patches P_x, P_y to be *structurally equivalent* $P_x \sim P_y$, if there exists a bijective function $g : \mathcal{I} \to \mathcal{I}$ such that

$$\forall z \in \mathcal{N}_x : P_x(z) = g(P_y(z)).$$

For an illustration, the first two patches in Fig. 2 are structurally equivalent, in contrast to the third one. The requirements f and f' have to fulfill are

(**R1**) **Locality preservation:**

$$\|P_x - P_y\| < \varepsilon \Rightarrow \|f(P_x) - f(P_y)\| < \varepsilon'. \tag{4}$$

(**R2**) **Structural equivalence:**

$$P_x \sim Q_x \iff f(P_x) = f'(Q_x), \tag{5}$$

with reasonable ε and ε' depending on the chosen norm. The motivation behind the first property is to ensure that similar patches are mapped to similar descriptors, which is important for the robustness to noise and the capture range of the registration. The second property states that descriptors are identical, if and only if, the patches are structurally equivalent. This ensures, on the one hand '⇒', the desired structural representation, and on the other hand '⇐', a perfect discrimination of patches. Generally, the discrimination for a dense descriptor is less critical than for a sparse descriptor.

Note that in (R1) only patches from the same image are compared, because the calculation of the norm $\|\cdot\|$ between patches from different modalities is not meaningful. In contrast, we require the structural equivalence (R2) only for patches of different images, because no intensity mapping is required in the same image. This model is no longer satisfiable by a global function f, so that we have to employ a local function for each modality, indicated with f and f'.

### 4.1. Structural equivalence vs. modality invariance

We would like to point out that the structural representation is different to a *modality invariant* representation, which would be the final goal. 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 MR images in Fig. 5, whose appearance is significantly different. It is also clear that structural 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 structural images, but is rooted in the multi-modal registration scenario, and therefore also affects multi-modal measures like MI. The application of robust metrics for comparing structural images, such as the robust M-estimation (Li et al., 1998), can limit the influence of those outliers.

5. Entropy images

A possible interpretation of the similarity between images is to consider whether intensity changes occur at the same locations. An example to quantify the intensity change is the calculation of the
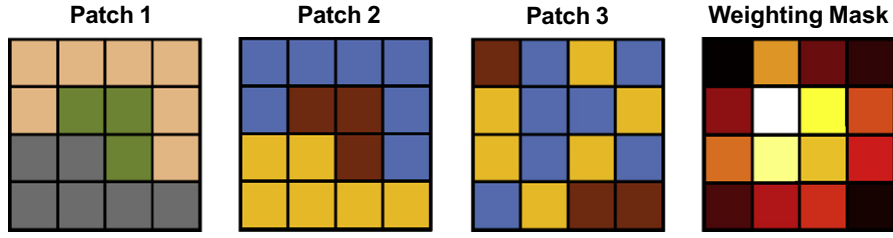
Fig. 2. Patches 1 and 2 show the same structure but encoded with different intensities. All three patches have the same entropy of 2.0749 using an identical spatial weighting. Integrating a location dependent weighting with the modified Gaussian weighting mask shown on the right, we can differentiate patch 3 from the others.
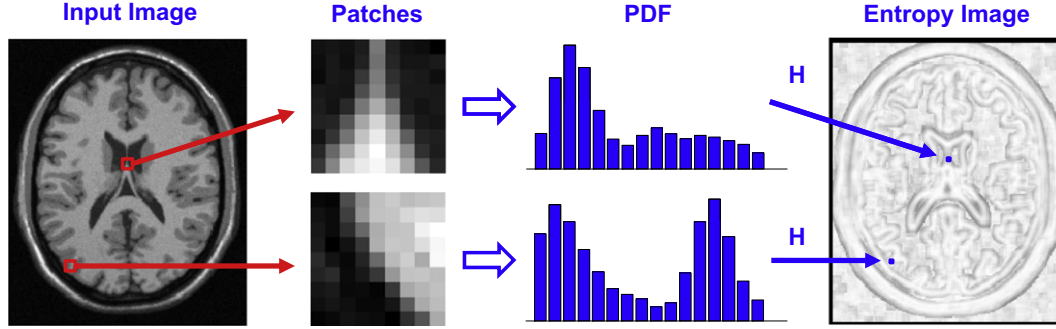
Fig. 3. Illustration of the process for calculating entropy images. For each pixel in the image, the local neighborhood patch is selected. For these patches, the PDF of the intensity is estimated, in this example with the Parzen-window method. Finally, the entropy is computed and the result is stored in the corresponding location in order to create the entropy image.

image gradient (Haber and Modersitzki, 2007). This is, however, not suitable for a structural representation because of its dependency on the similarity values. A more general concept is to quantify the information content 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

$$H(Y) = -\sum_{i \in \mathcal{I}} p(Y = i) \cdot \log p(Y = i), \tag{6}$$

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

$$D_x^I = H(I|_{\mathcal{N}_x}). \tag{7}$$

The construction of entropy images is illustrated in Fig. 3.

5.1. Verification of structural properties

In this section, we verify whether the entropy images fulfill the theoretical requirements on a structural representation. Since the entropy is calculated on images from both modalities, we do not have a local function f for every modality. We could therefore only fulfill (R1) and (R2) if we relax the requirements and allow patches P_x, Q_x to be from I as well as J. Verifying these relaxed requirements, we see that (R1) is fulfilled because small changes in the patches also lead to small changes in the entropy, because of the smoothness. The structural equivalence, "⇒" of (R2), is also fulfilled because the value of the entropy is invariant to the permutation of the bins in a histogram, which is the effect of the intensity mapping g.

In order to be able to assess the discrimination ability of the descriptors, "⇐" of (R2), we quantify the number of structurally different patches. Let $\lambda = |\mathcal{I}|$ be the number of intensity levels, and $k = |\mathcal{N}_x|$ be the cardinality of the patch. We assume $\lambda \geqslant k$, with typical values for $\lambda = 256$ and $k = 10 \times 10$. For these numbers, we exemplarily indicate the order of magnitude of the number of

patches in the subsequent analysis. The total number of different patches η_1 is calculated with

$$\eta_1 = \lambda^k \approx 10^{240}. \tag{8}$$

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

$$\eta_2 = B(k) = \frac{1}{e} \sum_{l=0}^{\infty} \binom{l^k}{l!} \approx 10^{115}. \tag{9}$$

This corresponds to the number of equivalence classes of the structural equivalence relation $\sim$. Patch 1 and 2 in Fig. 2 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. From a practical point of view, however, it would require more than 47 bytes per pixel to store up to 10^{115} different values, which could exceed the memory limit for volumetric data and decelerate the registration.

The number of different distributions is

$$\eta_3 = \binom{\lambda + k - 1}{k} \approx 10^{90}, \tag{10}$$

which corresponds to ball picking of unordered samples with replacement.

In order to make distributions fulfill the structural equivalence, we have to sort the entries of the distribution
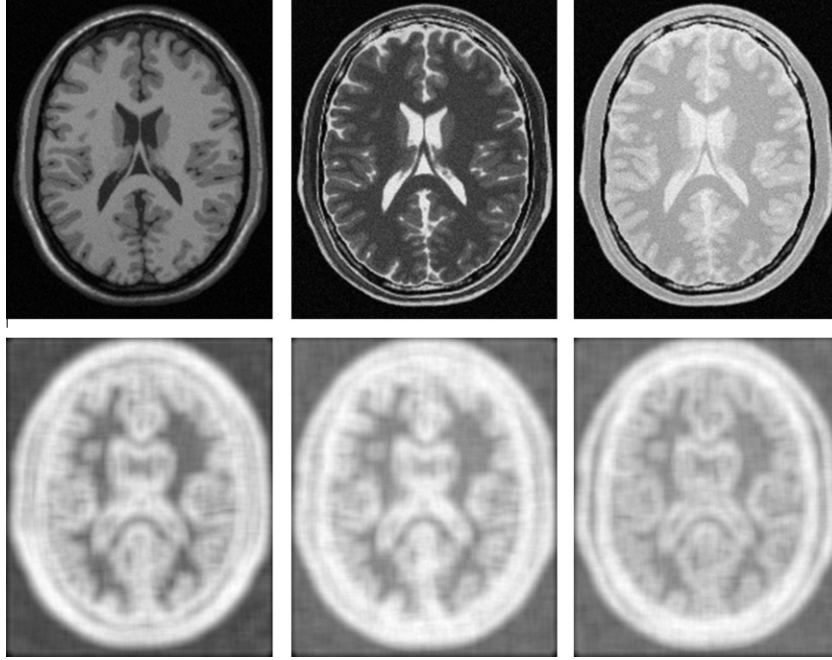
$$p' = \text{sort}(p). \tag{11}$$
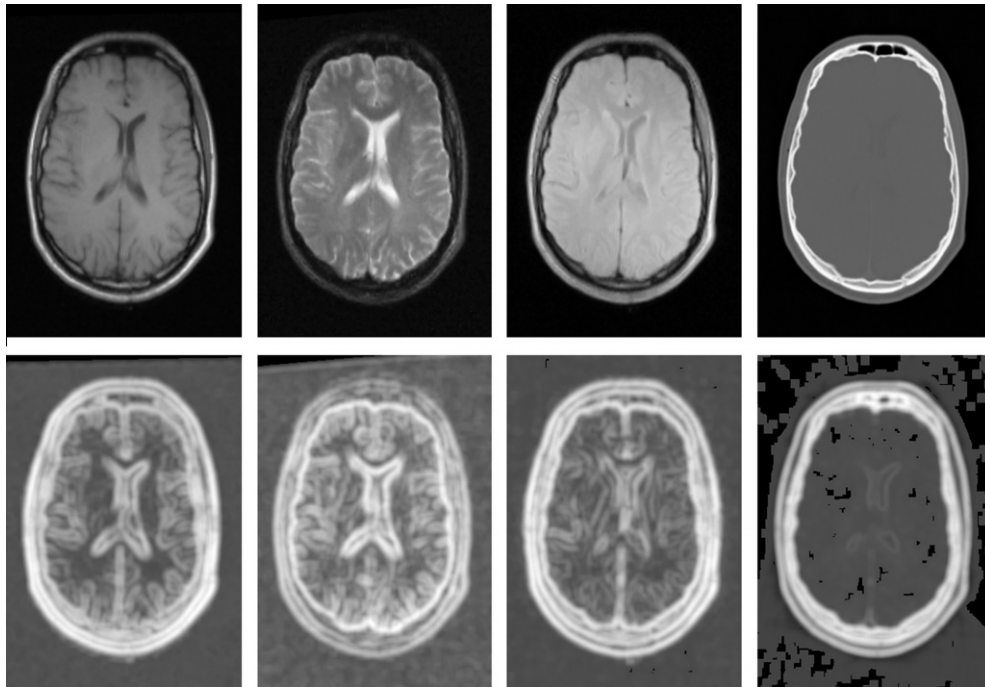
The number of sorted distributions p' is

$$\eta_4 = \mathcal{P}(k) \approx \frac{1}{4k\sqrt{3}} e^{\pi\sqrt{2k/3}} \approx 10^8 \tag{12}$$

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

Fig. 4. Multi-modal images (T1, T2, PD) from BrainWeb dataset together with entropy images used for rigid registration.
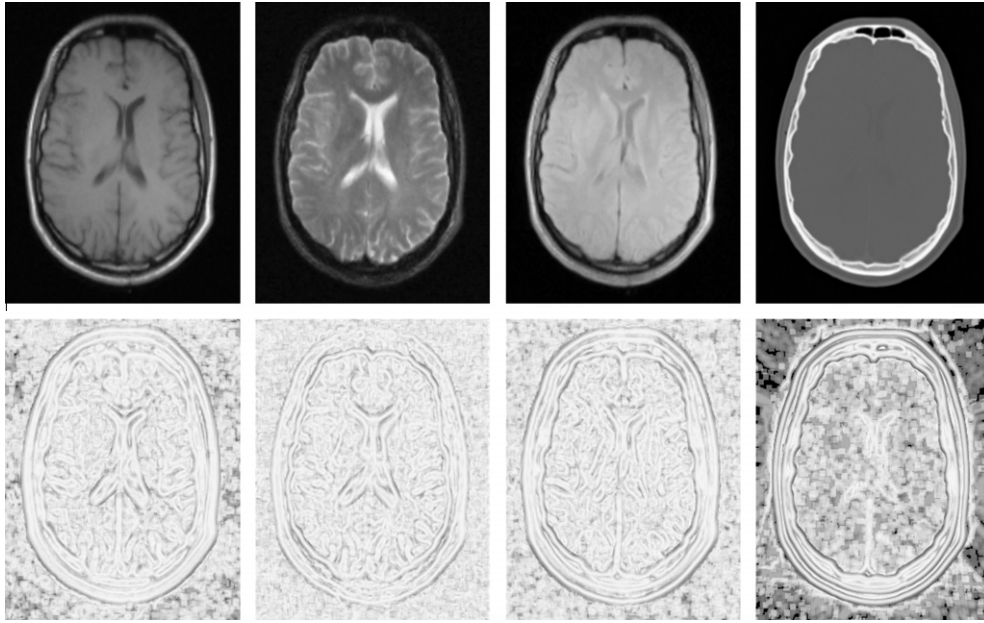

Fig. 5. Images from RIRE dataset (T1, T2, PD, CT) together with entropy images used for rigid registration.

The final step, the mapping from ordered histograms to real values, is performed with the entropy formula in Eq. (6). For $k = 2$, the entropy uniquely assigns each ordered histogram a scalar. However, for $k \geqslant 3$ the mapping is no longer injective and consequently η_4 presents an upper bound to the number of different entropy values for patches. Generally, the desired number η_2 is much higher than the maximally achievable η_4, so that "⇐" of property (R2) is not fulfilled. Although the discrimination is more critical for a sparse than for a dense descriptor, we try to improve it by adding a spatial weighting in the density estimation in Section 5.2.2.

5.2. Details on entropy estimation

There are several processing steps in the entropy estimation that influence the appearance of the entropy image, compare Figs.

Fig. 6. Multi-modal images (T1, T2, PD, CT) from RIRE dataset together with entropy images used for deformable registration.
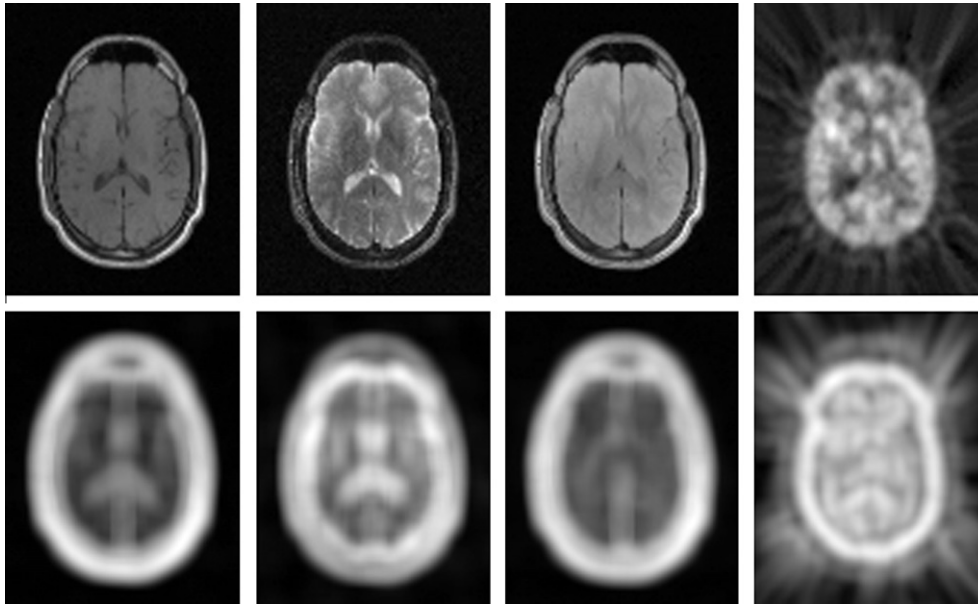
Fig. 7. Images from RIRE dataset with PET image (PET of lower resolution, 128 × 128) and corresponding entropy images.

5 and 6, where different parameters are used to create the entropy images. We evaluate the effect of those parameters for rigid and deformable registration. We present details about the experimental setup of the evaluation in Section 7.

5.2.1. Local neighborhood

The size of the local neighborhood is important because it determines the cardinality of the samples for the density estimation. Consequently, there is a trade-off between a small local neighborhood, in order to keep the estimation local, and a large neighborhood, to have sufficient statistics for an accurate estimation. For rigid registration, we evaluate local neighborhoods ranging from 5×5 to 21×21 pixels, where we found patches of size 11×11 to be a good compromise. For deformable registration, we prefer smaller patch sizes to have a more local description,

and therefore test them in the range from 5×5 to 13×13, with good results for 7×7 patches. In 3D, we achieved good results with $9 \times 9 \times 9$ patches, where we adapt the neighborhood accordingly for anisotropic spacing.

An interesting effect of changing neighborhood sizes is the smoothness of the entropy images. We show images that are created with patch sizes ranging from 5×5 to 21×21 in Fig. 8. We observe smoother entropy images for larger patch sizes. This is useful for registration, because a common approach is to perform registration on several resolution and smoothness levels. However, instead of filtering with a Gaussian kernel, we change the neighborhood size. The similarity graphs for various neighborhood sizes for T1-PD registration in Fig. 8 show wider cost functions for larger patch sizes, as we would expect them for Gaussian filtering with different variances.

(a) 21 × 21 (red)   (b) 15 × 15 (blue)   (c) 11 × 11 (green)   (d) 7 × 7 (cyan)   (e) 5 × 5 (magenta)   (f) Similarity plot
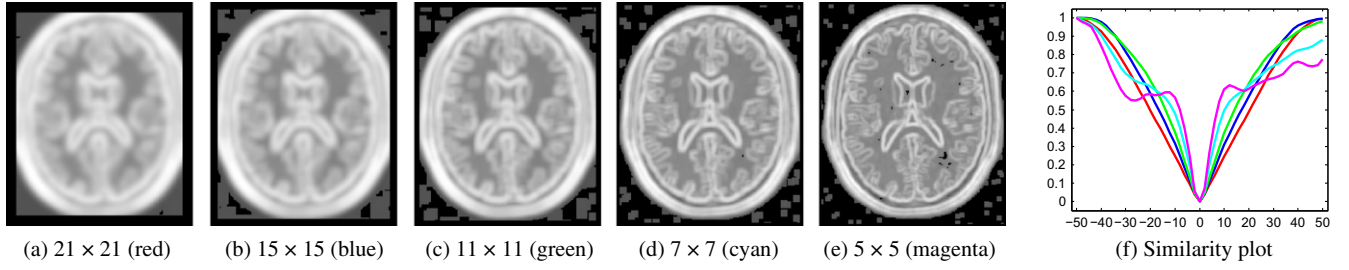
Fig. 8. Entropy images for varying patch sizes of BrainWeb T1. Similarity plot for rotation between entropy images of T1 and PD. Color of the curves as specified below the images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
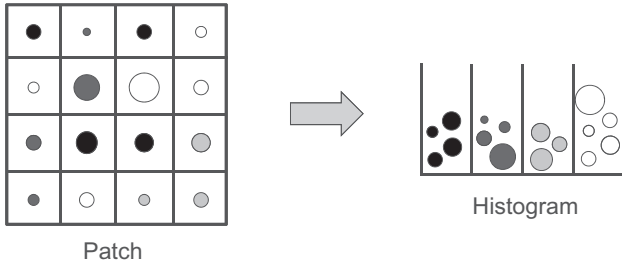
Fig. 9. The spatially varying weight for the histogram calculation is illustrated by the size of the balls. The discrimination of patches is improved, because two patches with the same intensity values, but at different locations, may lead to different histograms. The discrimination is dependent on the selected spatial weighting.

5.2.2. Spatially-weighted density 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 η_2 and η_4. Consider, for instance, the three patches in Fig. 2. While it is desirable 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, so that all three patches have the same entropy. 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 has the same contribution to the respective bin. We introduce a spatial weighting function $\omega : \mathcal{N}_x \to \mathbb{R}$, assigning a weight to each patch location. The histogram update changes to

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

This is illustrated in Fig. 9. We obtain the Parzen-window density estimation by convolution with a Gaussian kernel and corresponding normalization.

In our experiments, we use a Gaussian, a modified Gaussian, and the identity as weighting functions ω. The identity corresponds to the usual density estimation. For the Gaussian we set $\omega(y) = G_\sigma(\|y - c\|)$ with c the patch center, see Fig. 10 for an illustration. The discrimination between patches is not optimal because the Gaussian is rotational symmetric. We therefore modify the Gaussian weighting function, see Fig. 2, 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
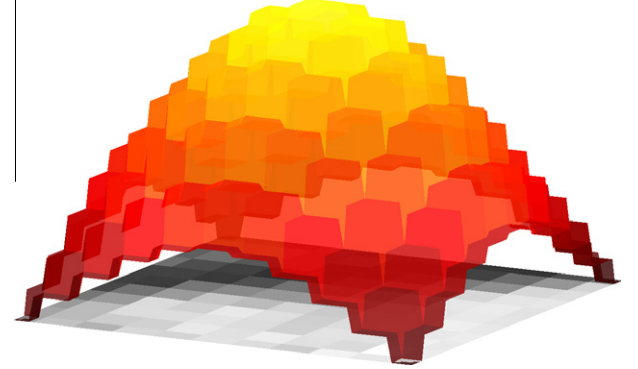
Fig. 10. Image patch (gray) with the corresponding spatial weighting (color) used for the histogram calculation. Illustrated is a Gaussian weighting. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

weighting mask, so that they are unique with respect to addition and that the dynamic range is minimal, is to use $2^1, 2^2, \ldots, 2^{|\mathcal{N}_x|}$. 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.

5.2.3. 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

$$\min_x = \inf_{y \in \Omega} I(y) \quad \max_x = \sup_{y \in \Omega} I(y) \tag{14}$$

or the local extrema in the patch

$$\min_x = \inf_{y \in \mathcal{N}_x} I(y) \quad \max_x = \sup_{y \in \mathcal{N}_x} I(y). \tag{15}$$

Our experiments confirm that a global approach is better suited for rigid, while the local approach is better suited for deformable registration. We show entropy images with the global approach in Figs. 4, 5, and 7 whereas the local approach is illustrated in Fig. 6.

5.2.4. Entropy measure

The Shannon entropy is one in a group of measures to calculate the entropy of a random variable, so that we evaluate its influence on registration.

A whole class of entropy measures is deduced from the Rényi entropy (Rényi, 1961)

$$H_\alpha(Y) = \frac{1}{1 - \alpha} \log \left(\sum_{i \in \mathcal{I}} p(Y = i)^\alpha \right), \tag{16}$$

defined for $\alpha \geqslant 0$ and $\alpha \neq 1$. For $\alpha \to 1$, it converges to the Shannon entropy, $H_1 = H$. Commonly used in the group of Rényi entropies is

the case for $\alpha = 2$. An alternative entropy measure, not in the group of Rényi measures is the Burg entropy (Mansoury and Pasha, 2008)

$$H_{Burg}(Y) = \sum_{i \in \mathcal{I}} \log p(Y = i). \tag{17}$$

6. Laplacian images

As an alternative structural representation, we propose the application of manifold learning. Laplacian eigenmaps present an optimal solution to one of the requirements for a structural representation. This is the preservation of locality, meaning that patches that are close in high-dimensional patch space are mapped to a close structural representation. The requirement of structural equivalence across modalities is guaranteed by the comparable internal similarity.

Manifold learning is an approach applied for non-linear dimensionality reduction and data representation. The task of dimensionality reduction is to find the underlying structure in a large set of points embedded in a high-dimensional space and to map these points to a low-dimensional space preserving the structure. Manifold learning has recently gained much attention to assist image processing tasks such as segmentation (Zhang et al., 2006), registration (Rohde et al., 2008; Hamm et al., 2009), tracking (Lee and Elgammal, 2007; Wachinger et al., 2010b), recognition (Arandjelovic and Cipolla, 2007; Wachinger et al., 2010a), and computational anatomy (Gerber et al., 2009). Common techniques for manifold learning are Isomap (Tenenbaum et al., 2000), local linear embedding (Roweis and Saul, 2000), and Laplacian eigenmaps (Belkin and Niyogi, 2003). We focus on Laplacian eigenmaps because the technique is well founded on mathematical concepts (Laplace Beltrami operator) and computationally efficient.

An overview of the calculation of Laplacian images is presented in Fig. 11. We use all the patches in an image to build a neighborhood graph, approximating the manifold embedded in high dimensional patch space. Subsequently, the graph Laplacian is calculated to find an optimal mapping to low-dimensional space. Since the

embedding in low-dimensional space is arbitrary, as long as it preserves the locality, we have to align embeddings from different modalities with an affine point-based registration. This finally leads to the structural representation that is used in the intensity-based registration

6.1. Laplacian eigenmaps

Considering k points $\mathbf{a}_1, \ldots, \mathbf{a}_k$ in $\mathbb{R}^N$ lying on a manifold $\mathcal{M}$, we want to find a set of corresponding points $\mathbf{b}_1, \ldots, \mathbf{b}_k$ in the low-dimensional space $\mathbb{R}^n (n \ll N)$. We assume a twice differentiable function $m : \mathcal{M} \to \mathbb{R}^n$, $\mathbf{a}_i \mapsto \mathbf{b}_i$. It is shown in Belkin and Niyogi (2003) that the following holds for the relationship between the distances on the manifold and embedding space

$$\|m(\mathbf{a}_i) - m(\mathbf{a}_j)\| \leqslant d_{\mathcal{M}} \cdot \|\nabla m(\mathbf{a}_i)\| + o(d_{\mathcal{M}}), \tag{18}$$

with $\mathbf{a}_i, \mathbf{a}_j \in \mathcal{M}$ and $d_{\mathcal{M}} = \text{dist}_{\mathcal{M}}(\mathbf{a}_i, \mathbf{a}_j)$ the manifold distance. We see that $\|\nabla m(\mathbf{a}_i)\|$ indicates how close nearby points are mapped. Consequently, a map that best preserves locality on average is found with the following minimization

$$\arg \min_{\|m\|_{L^2(\mathcal{M})} = 1} \int_{\mathcal{M}} \|\nabla m(\mathbf{a}_i)\|^2 \, d\mathbf{a}_i. \tag{19}$$

It is this optimization problem, for which Laplacian eigenmaps provide an optimal solution, by calculating eigenfunctions of the Laplace Beltrami operator.

6.2. Verification of structural properties

The optimal locality preservation as it is provided by the Laplacian eigenmaps, see Eq. (19), is exactly what was required for the structural representation in (R1). We only have to identify the points $\mathbf{a}_i$, $\mathbf{a}_j$ with the patches P_x, P_y. The mapping m is therefore a suitable candidate for the function that provides the structural representation f.

For the second property (R2), we consider manifolds $\mathcal{M}$ and $\mathcal{M}'$ for two different modalities with patches $P_x \in \mathcal{M}$ and $Q_x \in \mathcal{M}'$. The
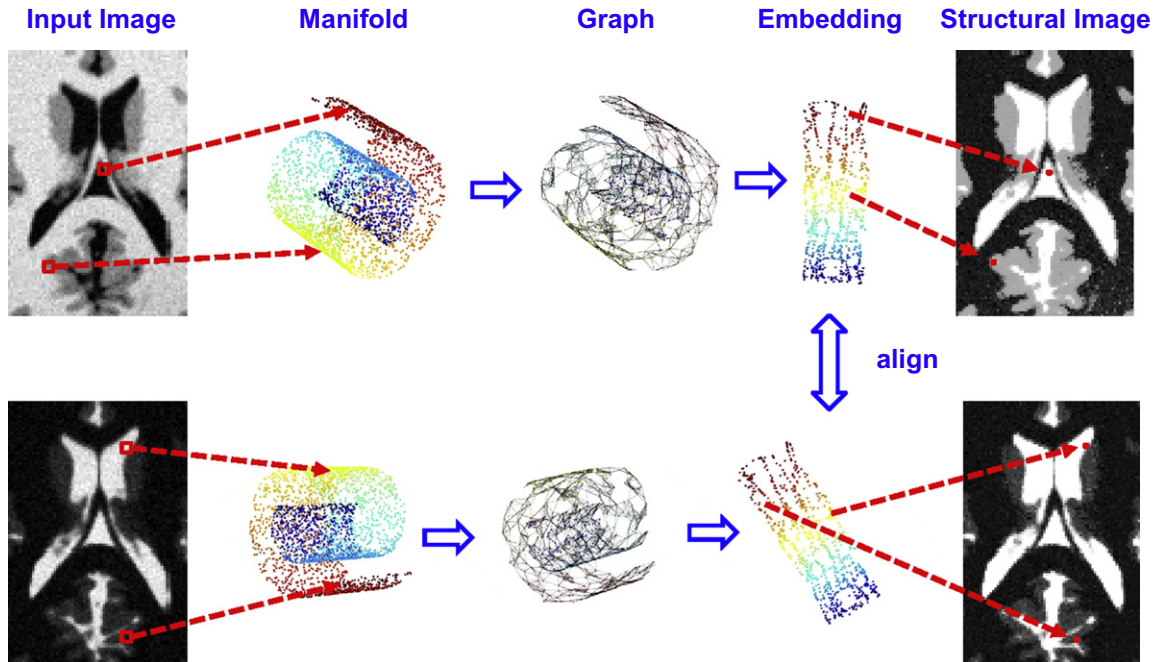
Fig. 11. Structural representation with Laplacian eigenmaps. Patches of images lie on a manifold in high-dimensional patch space. The manifold is approximated by the neighborhood graph. The low-dimensional embedding is calculated with the graph Laplacian. Embeddings from different modalities have to be aligned to obtain the final representation.

mappings m and m' provide $m : P_x \mapsto p_x$ and $m' : Q_x \mapsto q_x$, with p_x and q_x being the low-dimensional representations. Since the intensity, with which objects are depicted in the images, varies with the modality, the two manifolds $\mathcal{M}$ and $\mathcal{M}'$ are not directly comparable. Considering, however, the assumption that the internal similarity of both modalities is equivalent, as in Penney et al. (2008), we conclude that the structure or shape of both manifolds is similar. Since Laplacian eigenmaps preserve locality when embedding the manifold in a low-dimensional space, this structure is preserved in low dimensions. We could then directly use the coordinates of p_x as descriptor for the corresponding location D_x. This is, however, not possible because the embedding of the structure in low-dimensional space is arbitrary, as long as it preserves the locality. The embeddings of both manifolds $\mathcal{M}$ and $\mathcal{M}'$ are therefore only similar when correcting for rotation, translation, and scale. Consequently, an affine registration between the point sets $\mathcal{P} = \{ p_x : x \in \Omega \}$ and $\mathcal{Q} = \{ q_x : x \in \Omega \}$ has to be performed. Recently, an affine ICP for point set alignment was proposed in Du et al. (2010) that performs a fast registration. The coordinates of the registered embeddings finally provide the structural descriptors.

We conclude that m fulfills the requirements (R1) and (R2) for a structural representation. It has therefore theoretical advantages in comparison to entropy images, since they only fulfill the relaxed properties, and further, the preservation of locality is optimal for m.

6.3. Application of Laplacian eigenmaps

In this section, we describe the application of Laplacian eigenmaps to our problem in more details. We consider one dimension of the ambient space for each image pixel of the patches. We use patches of size 15×15, so that $N = 225$. This size proved to be a good compromise between too small patches that do not contain enough structural information, and too large patches that contradict the required locality and further lead to a higher computational burden.

We construct a graph with a node for each point P_x and with edges connecting neighboring nodes. The neighborhood can be defined with an δ-neighborhood around each point, so P_y is in the δ-neighborhood of P_x if $\| P_x - P_y \|^2 < \delta$, with $\| \cdot \|$ the Euclidean norm. Although this is geometrically motivated, a disadvantage is the selection of the parameter δ. In our implementation, we search instead for the $l = 500$ nearest neighbors and add edges between them in the adjacency graph. Further, heat kernel-based weights are assigned to the edges with $w_{xy} = e^{-\| P_x - P_y \|^2 / (2 \cdot \sigma^2)}$ and σ^2 the variance.

We select $n = 1$ as dimension for the low-dimensional space. The reasons for not increasing this value are, first, that we obtain good results, and second, that for $n > 1$ we would have to store a vector in each pixel position instead of a scalar. This increases the computational complexity and memory consumption of the registration, and moreover, makes the visualization more challenging. Additionally, the low-dimensionality facilitates the affine alignment to a correction of scale and shift. In fact, a robust normalization accounting for outliers and flipping, makes the point-based registration in this case even superfluous. This holds the advantage of making the Laplacian images more comparable to entropy images, because they, as well, do not need any intermediary step.

Although the Laplacian images have superior theoretical properties than entropy images, this comes at a much higher computational cost. For the construction of the neighborhood graph, all patches have to be compared among each other. This means that the computational complexity increases quadratically with the number of patches. Thinking about its application for the alignment of volumetric data, this becomes even more challenging for the comparison of all subvolumes. This is different for entropy images, where the complexity increases only linearly. The Laplacian images present a global approach to the creation of structural images, while the entropy images work entirely locally.

7. Experiments with entropy images

For entropy images, we conduct experiments on T1, T2, and PD-weighted MR images from the BrainWeb database[1] and CT, T1, T2, PD, and PET (Positron Emission Tomography) images from the Retrospective Image Registration Evaluation (RIRE) database.[2] We work with BrainWeb images containing 3% noise and 20% intensity non-uniformity, in order to achieve realistic results. For both databases the ground truth alignment is provided. We depict axial slices of the original and entropy images in Figs. 4–7. The average time for the creation of an entropy image in C++ is 0.078 s for a slice of 256×256, and 11.01 s for a volume of $512 \times 512 \times 29$.

For our deformable experiments we use the freely available deformable registration software DROP,[3] 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 guarantees that similar results are obtainable with alternative registration approaches.

7.1. Rigid registration

For rigid registration, the standard configuration for the entropy images is: 11×11 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 combinations of both datasets. In Fig. 12 we show an excerpt, with a complete list of graphs in the Supplementary material. The plots are created by rotating the images around the image center for the range of $[-40°, 40°]$, with 0° 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 ±20 mm in translation and ±20° 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 Tables 1 and 2, we show the absolute mean error for each pose parameter and the total mean error for translation t_{total} and rotation r_{total} for the two datasets. Additionally, we indicate the standard deviation of the error for translation t_{std} and rotation r_{std}. On the BrainWeb dataset eSSD and MI lead to comparable results. The large values of the standard deviation for MI indicate that some of the registrations did not converge correctly. Compared to the low standard deviations for eSSD, we conclude that the registration is more robust with eSSD. On the MR volumes of the RIRE dataset, eSSD performs significantly better than MI, matching with our observations from the similarity plots. For the alignment with CT volumes the registration error increases

---

(a) T1-T2

(b) T1-PD

(c) CT-T2

(d) PET-T2
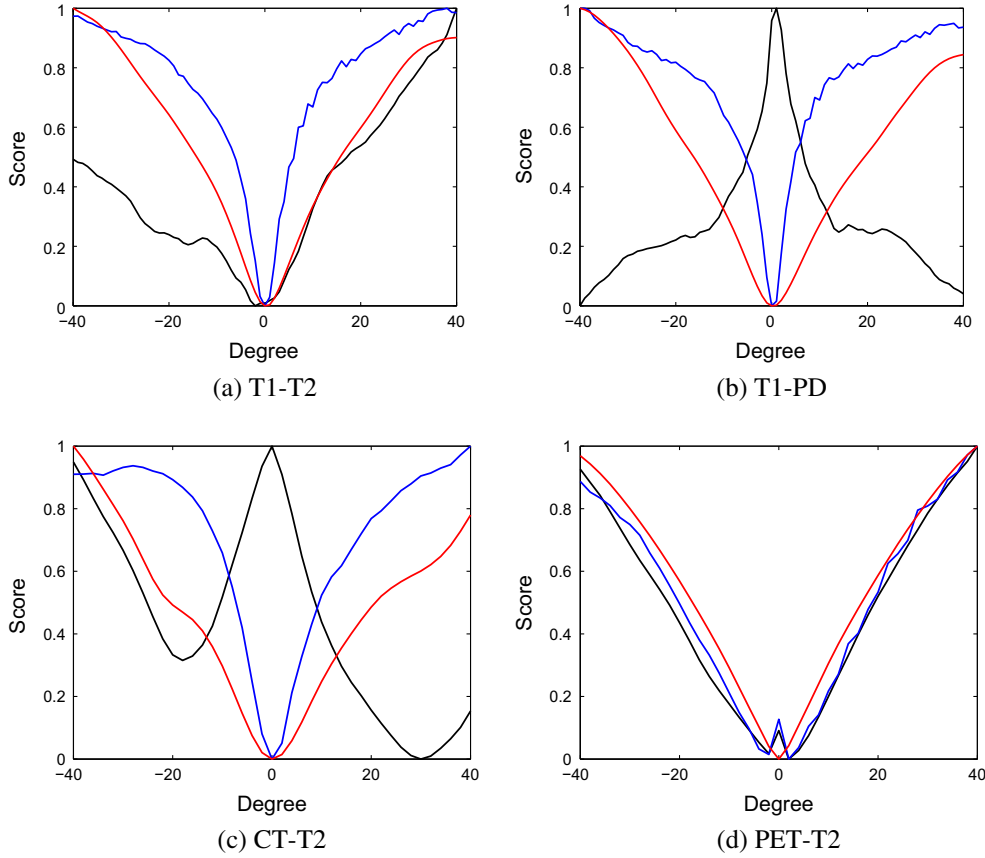
Fig. 12. Similarity plots for RIRE dataset (black: SSD, blue: MI, red: eSSD). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Pairwise random registration study for BrainWeb with translational error t in mm, rotational error r in degree.

Data set	Sim	t_x	t_y	t_z	r_x	r_y	r_z	t_{std}	r_{std}	t_{total}	r_{total}
T1-T2	MI	0.189	0.247	0.162	0.430	0.331	0.122	1.570	2.849	**0.376**	0.579
T1-T2	eSSD	0.165	0.209	0.622	0.051	0.056	0.282	0.077	0.170	0.695	**0.317**
T1-PD	MI	0.251	0.319	0.257	0.359	0.417	0.165	1.906	2.303	**0.506**	0.621
T1-PD	eSSD	0.155	0.119	0.538	0.047	0.064	0.334	0.077	0.152	0.5929	**0.365**
T2-PD	MI	0.171	0.112	0.143	0.192	0.156	0.119	0.792	0.861	0.274	0.290
T2-PD	eSSD	0.046	0.041	0.141	0.045	0.040	0.044	0.039	0.025	**0.161**	**0.085**

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. The good registration results for eSSD on the RIRE data are confirmed by the low standard deviations. Comparing the errors for the different pose parameters, we observe that the translational error along the z-axis t_z is commonly larger than along the other axes. This is due to the lower resolution along the z-axis. The low resolution is also the reason for the lower rotational error around the z-axis r_z in comparison to r_x and r_y.

We measure an average computation time for the evaluation of SSD being a factor of 15 faster than MI. This leads to a quick amortization of the additional processing time needed for the creation of the entropy images.

7.2. Deformable registration

We first evaluate the application of gradient images (Haber and Modersitzki, 2007) for deformable registration. In Fig. 13 we show

two synthetic images that model a possible multi-modal scenario, with image regions being depicted in different colors. We calculate the gradient fields of both images and overlay them, once for the upper and once for the lower T-junction. Further, we smooth the synthetic images to create a more realistic scenario and show the gradient fields again. We observe that the gradients are pointing in different directions, where the smoothing propagates this effect to a larger region. Contrary, the entropy images consistently 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 Figs. 14 and 15, and Table 3. The error of the registration with the original images using normalized mutual information is stated in the table as

Table 2
Pairwise random registration study for RIRE with translational error t in mm, rotational error r in degree. The bold values indicate the results with lower registration error.

Data set	Sim	t_x	t_y	t_z	r_x	r_y	r_z	t_{std}	r_{std}	t_{total}	r_{total}
T1–T2	MI	0.719	0.395	1.531	1.594	2.252	1.467	0.235	0.177	1.754	3.139
T1–T2	eSSD	0.042	0.224	0.396	1.120	1.582	0.538	0.046	0.057	**0.461**	**2.013**
T1–PD	MI	0.190	0.251	0.856	0.635	0.877	0.546	0.290	0.180	0.944	1.231
T1–PD	eSSD	0.061	0.048	0.431	0.344	0.648	0.558	0.050	0.048	**0.442**	**0.926**
T2–PD	MI	0.196	0.344	2.068	2.365	1.196	0.952	0.718	2.186	2.141	2.976
T2–PD	eSSD	0.093	0.110	1.065	1.984	1.196	0.308	0.279	0.493	**1.080**	**2.355**
CT–T1	MI	1.925	1.004	1.312	1.718	2.951	0.763	2.394	5.529	2.710	3.951
CT–T1	eSSD	0.963	1.269	0.702	2.433	0.728	0.169	2.990	9.577	**1.997**	**3.089**
CT–T2	MI	4.567	1.501	2.314	3.896	9.738	1.332	4.543	9.783	**5.806**	**11.634**
CT–T2	eSSD	1.288	8.488	6.025	35.905	1.903	0.885	7.932	28.631	11.063	36.896
CT–PD	MI	2.348	0.613	1.394	1.395	4.056	0.805	2.677	5.622	2.943	4.784
CT–PD	eSSD	0.770	0.988	0.475	1.487	0.248	0.248	1.824	5.971	**1.442**	**1.619**
PET–T1	MI	9.071	7.730	13.409	29.226	23.578	4.945	13.373	22.343	20.869	46.234
PET–T1	eSSD	0.053	0.057	0.089	0.040	0.038	0.042	0.051	0.023	**0.135**	**0.078**
PET–T2	MI	9.027	8.566	11.631	35.444	22.239	5.522	11.545	20.217	20.705	52.131
PET–T2	eSSD	0.043	0.038	0.057	0.040	0.040	0.042	0.034	0.024	**0.093**	**0.079**
PET–PD	MI	10.131	7.685	13.354	27.693	24.451	4.565	12.853	20.023	21.165	45.606
PET–PD	eSSD	0.048	0.056	0.194	0.043	0.056	0.043	0.065	0.033	**0.217**	**0.094**
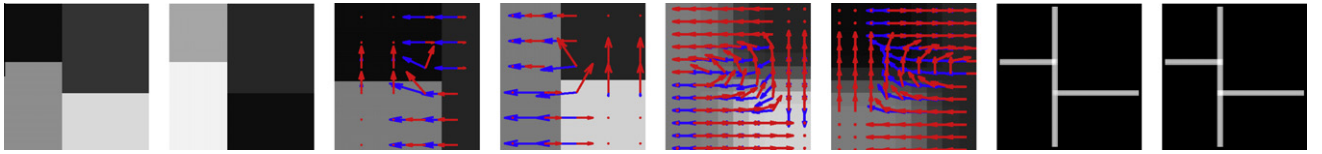
Fig. 13. Illustration of two synthetic multi-modal images (1st & 2nd) together with a zoom on gradient fields (3rd & 4th), gradient fields of smoothed images (5th & 6th) and entropy images (7th & 8th). The best structural representation is achieved with entropy images.
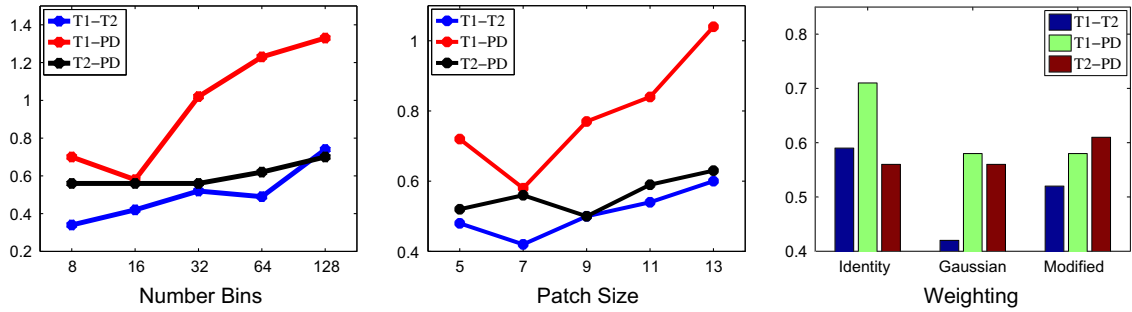
Fig. 14. Evaluation of error τ for deformable registration for varying bin number, patch size, and spatial weighting on Brainweb.
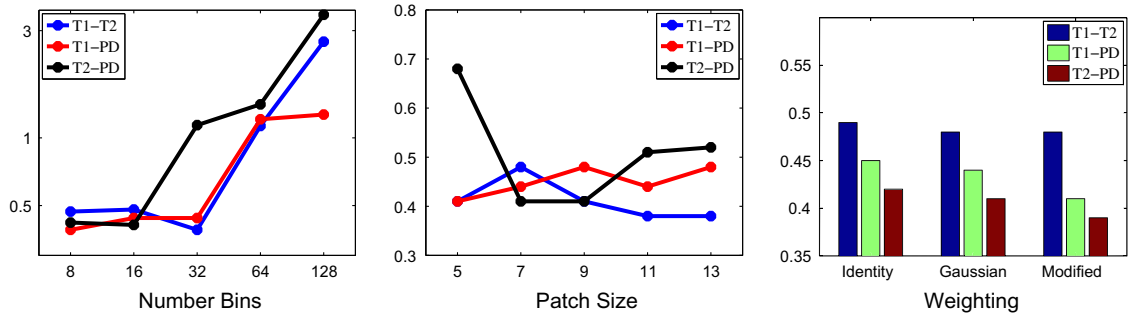
Fig. 15. Evaluation of error τ for deformable registration for varying bin number, patch size, and spatial weighting on RIRE.

reference. The standard configuration for the entropy image for deformable registration is: 7×7 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 Figs. 14 and 15, we see that best results are achieved around 16 bins. While reducing it further to 8 bins also leads to

Table 3
Registration errors τ in mm for various configurations for calculating the entropy images. (B: Brainweb, R: RIRE dataset).

| Technique	T1-T2$_B$	T1-PD$_B$	T2-PD$_B$	T1-T2$_R$	T1-PD$_R$	T2-PD$_R$	CT-T1$_R$	PET-T2$_R$
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
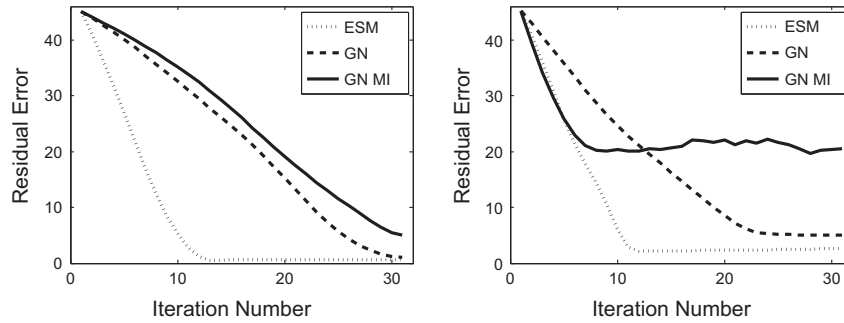
Fig. 16. Groupwise convergence study for BrainWeb (left) and RIRE (right) volumes.
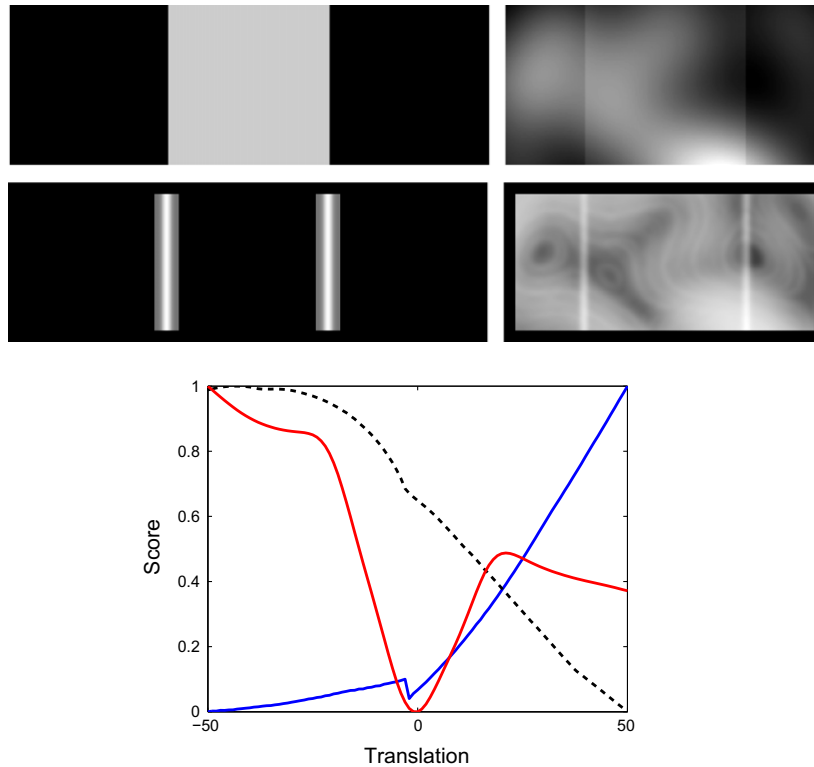
Fig. 17. First line, synthetic images with bias field (Myronenko and Song, 2009). Second line, entropy images. Third line, similarity plot with SSD (blue), MI (black, dashed), and eSSD (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

good results, increasing it further to 32 bins leads to an increase in error. A good compromise in the patch size for the different data-sets and modalities is approximately 7×7. 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 Table 3, we see that a local normalization of the intensity values and the Parzen-window method for the density estimation are essential for good deformable registration results. For the
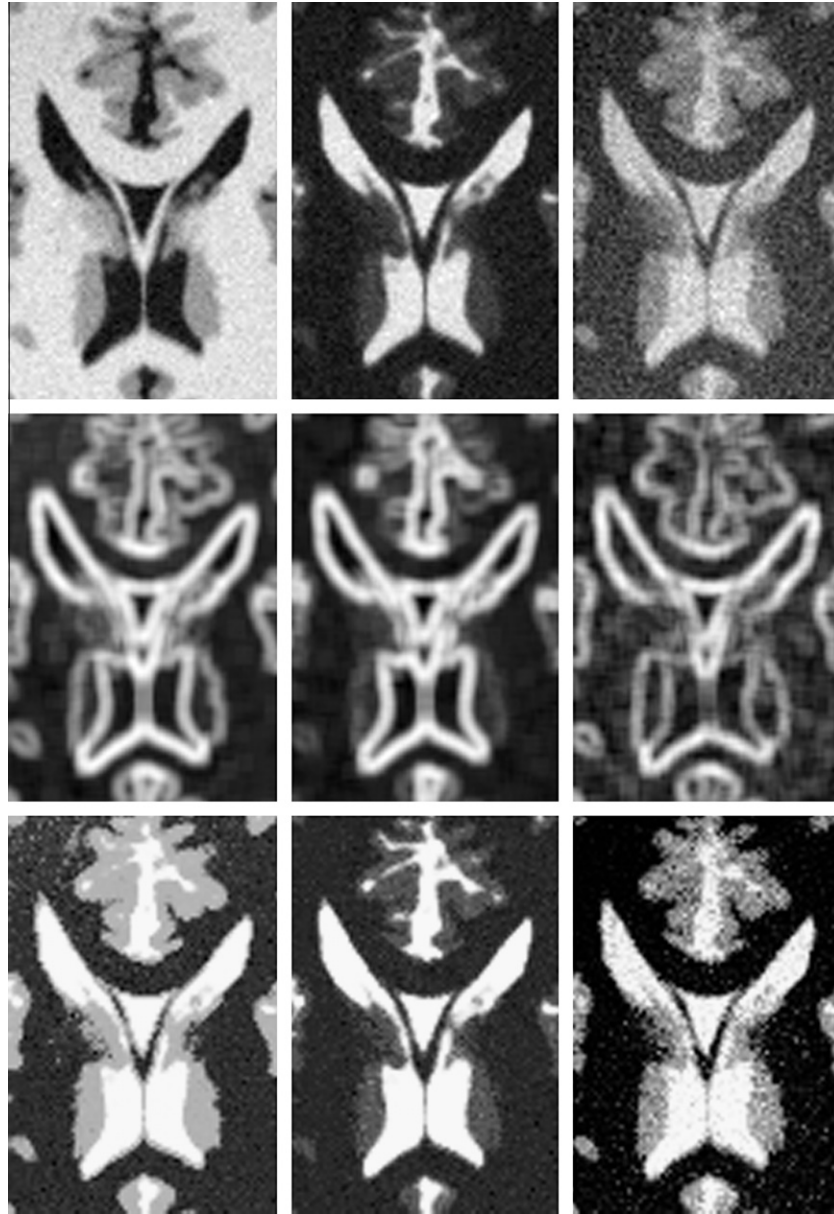
Fig. 18. T1 (left), T2 (middle), and PD (right) images. First line: original images. Second line: entropy images. Third line: Laplacian images.

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 on average 6.6 times faster than with NMI. This includes the time for the creation of the entropy images.

7.3. Groupwise registration

We perform a simultaneous, rigid registration study for the BrainWeb and RIRE 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 1 mm equal to 1°. The average residual error for each step is shown in Fig. 16. 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.

7.4. Bias field

A typical situation that challenges the application of mutual information is the registration of images, which contain high intensity non-uniformity due to the bias field. Recently, the residual complexity (Myronenko and Song, 2009) was presented as a technique that can cope with such situations. They motivated their approach with synthetic images, similar to those in Fig. 17. We
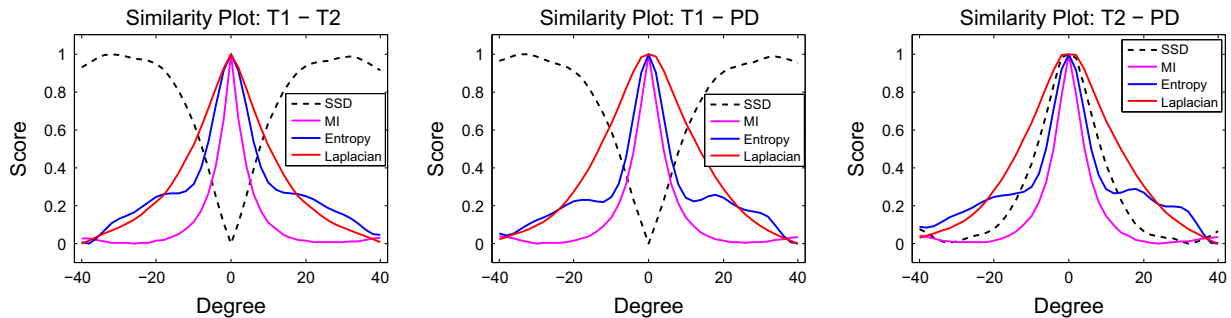
Fig. 19. Plot of similarity measures with respect to rotation of images in Fig. 18. Maxima indicate best alignment.
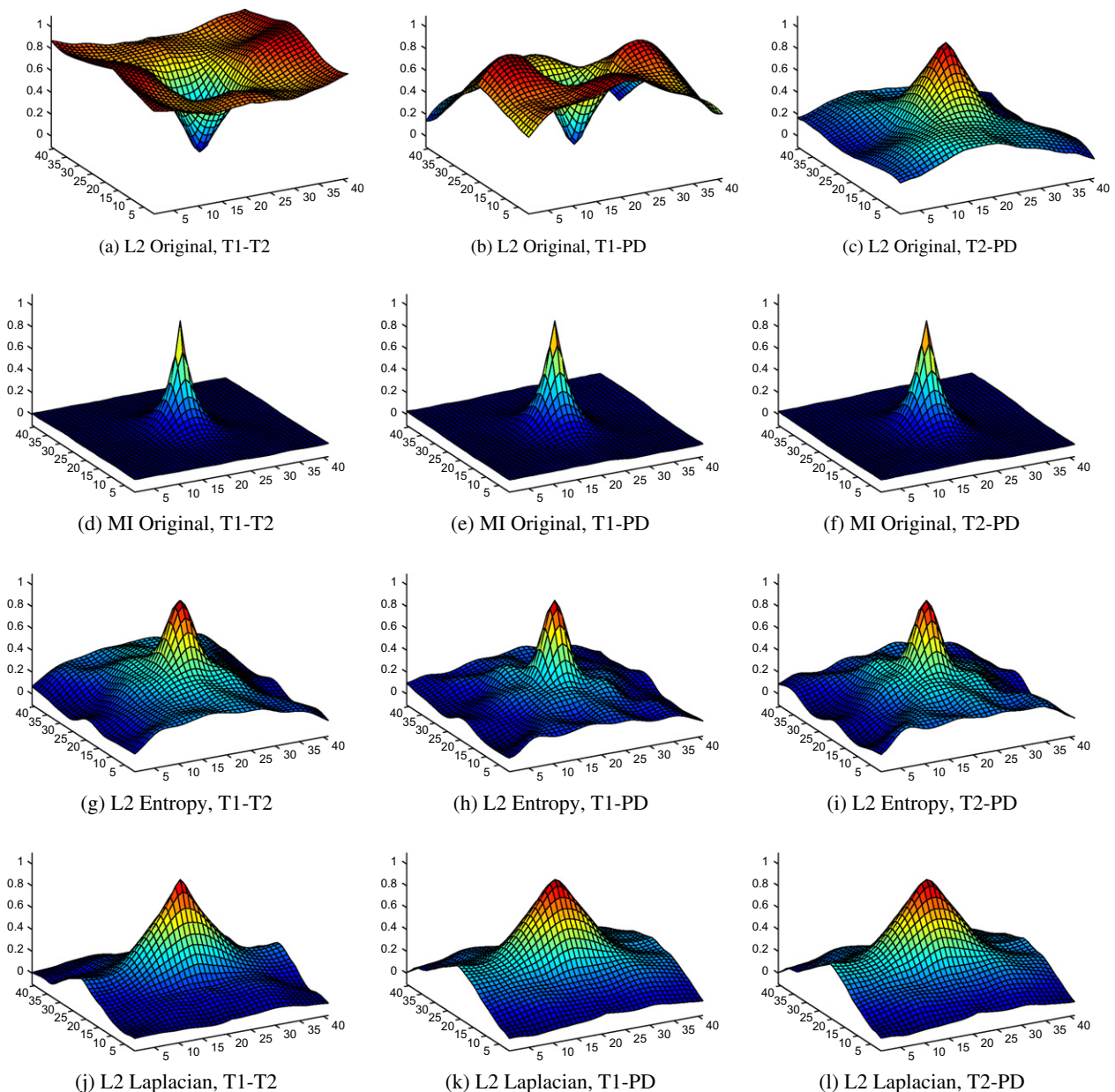

(a) L2 Original, T1-T2

(b) L2 Original, T1-PD

(c) L2 Original, T2-PD

(d) MI Original, T1-T2

(e) MI Original, T1-PD

(f) MI Original, T2-PD

(g) L2 Entropy, T1-T2

(h) L2 Entropy, T1-PD

(i) L2 Entropy, T2-PD

(j) L2 Laplacian, T1-T2

(k) L2 Laplacian, T1-PD

(l) L2 Laplacian, T2-PD

Fig. 20. Plot of similarity measures with respect to translation in *x* and *y* direction. Maxima indicate best alignment.

create plots of SSD and MI by translating the images, analogously to Myronenko and Song (2009). As expected, they fail in indicating the correct alignment. In contrast, eSSD correctly indicates the perfect alignment. We show a second set of images affected by bias

field in the Supplementary material with similar results. There is an interesting parallel, because it was shown that the usage of local statistics for the MI calculation (Yi and Soatto, 2009) is more robust to intensity non-uniformity. For the calculation of entropy
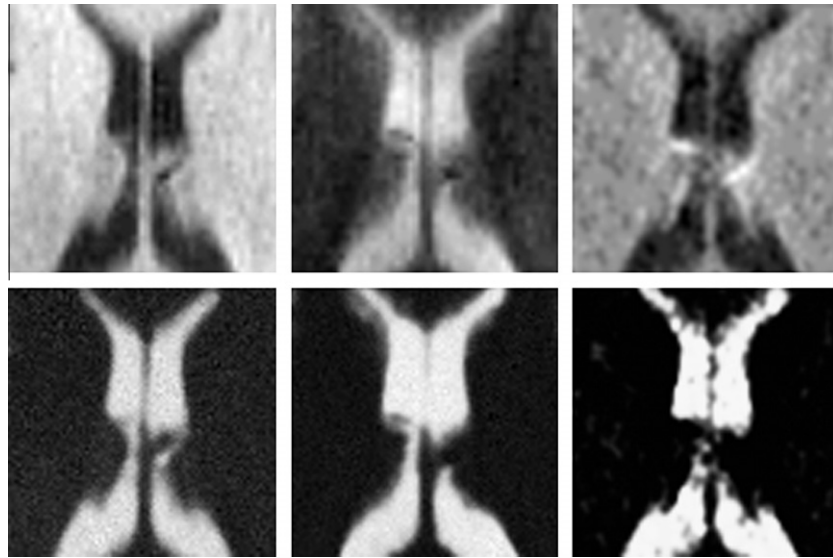
Fig. 21. T1 (left), T2 (middle), and CT (right) images. First line: original images. Second line: Laplacian images.

Table 4
Registration study for Laplacian images. Errors for translation t in mm and rotation r in degree. The bold values indicate the results with lower registration error.

Datasets	Similarity	r	t_x	t_y	RMS
T1-T2	L2	4.879	9.019	6.471	7.000
	MI	2.325	3.768	5.226	3.954
	Entropy	2.084	4.539	5.231	4.180
	Laplacian	2.584	2.061	2.168	**2.271**
T1-PD	L2	2.760	6.422	5.755	5.227
	MI	2.304	4.138	4.907	3.937
	Entropy	2.283	4.782	4.750	4.108
	Laplacian	1.750	3.007	1.929	**2.297**
T2-PD	L2	1.784	2.947	2.916	3.942
	MI	2.161	4.628	3.812	3.680
	Entropy	1.723	4.296	3.780	3.450
	Laplacian	1.171	2.350	1.984	**1.900**

images we also use local statistics. However, we have the advantage of estimating only marginal instead of joint distributions, which is more robust for few samples from small local neighborhoods.

8. Experiments with Laplacian images

We already mentioned in Section 6.3 that the computational complexity of the Laplacian images prohibits its practical usage on a large scale at the moment. We nevertheless want to validate if the theoretical advantages also lead to better registration results. We limit the analysis to 2D regions of the BrainWeb images (3% noise, 20% intensity non-uniformity) and RIRE images. In Fig. 18 we show the original images, the entropy images, and the Laplacian images. We can clearly observe the different nature of the entropy and Laplacian images. Entropy images resemble gradient images, pronouncing boundaries and changes in the images. The Laplacian images, however, look like the original images, but with a different coloring. We can further observe that the assumption of comparable internal similarities in the images is justified, because the appearance of the Laplacian images across the modalities is very similar. In Fig. 21 we show the results for the RIRE images with T1, T2, and CT. The assumption of comparable internal similarity between MR and CT brain is more challenging, because of the low contrast of internal brain structures in CT.

In order to quantify the promising visual appearance for image registration, we show surface plots of the similarity measures for translation and rotation in Figs. 19 and 20, respectively. We compare the usage of the L2 distance on the original images, MI on the original images, and L2 distance on entropy and Laplacian images for all combinations of multi-modal alignment. The maxima indicate the best alignment. MI shows a very sharp peak at the correct position, but seems to have a limited capture range. Entropy images also indicate the correct position, but the cost functions contain several local maxima. We observe the cost function with the largest capture range for the Laplacian images.

Although similarity plots give a good intuition about the performance of different similarity measures, it is only a registration study that shows the final quality. We perform a registration study for all multi-modal image combinations. The random starting position deviates up to ±15 mm in translation and ±15° in rotation from the correct pose. We show the average absolute error for translation and rotation, together with the overall root mean squared error (RMS), for 100 registration runs for each configuration in Table 4. We weight 1 mm equal to 1° to quantify translational and angular displacement from the ground truth in one single value. We see that the positive impression of the Laplacian images from the similarity plots is confirmed by the registration results. We obtain a significantly lower error in comparison to MI and entropy images. Our experiments therefore confirm the theoretical advantages of Laplacian images in comparison to entropy images in practice.

9. Discussion

We extensively evaluated the performance of entropy images for rigid, deformable, and groupwise registration. The results on the BrainWeb data are comparable to the application of MI. The registration results on the RIRE database are dependent on the modalities. We obtain very good results for the alignment of PET images, and in most other cases we are comparable or slightly better than MI. The application of ESM on entropy images did not cause any problems, and we obtained the expected faster convergence in comparison to Gauß–Newton. Finally, we indicated that entropy images are robust to intensity non-uniformity.

For the Laplacian images, we were able to validate that their superior theoretical properties also lead to better registration results. The computational complexity for the creation of these representations is, however, limiting. We think that the juxtaposition of entropy and Laplacian images is interesting because they present two completely different approaches for obtaining structural representations. While entropy images could be seen as non-linear filtering of an image, the Laplacian images identify self-similarities in an image. Entropy images present a local approach, and Laplacian images a global approach to the calculation of structural representations.

Thinking along the line of entropy images, one could imagine several alternative techniques to quantify the structure of a patch. We experimented, for instance, with the compressibility of patches or the calculation of the entropy from gradient instead of intensity histograms, with examples shown in the Supplementary material. These approaches, however, bear no obvious advantages for our applications and are more complex to compute. We are convinced that the simplicity of the entropy images together with the fast calculation and good performance leads to a high practical value. Further, there are many articles in the computer vision literature (Sizintsev et al., 2008; Wei and Tao, 2010) that deal with a fast local histogram calculation, such as integral histograms, distributive histograms, and the efficient histogram-based sliding window, that allow for a reduction of the processing time for certain configurations of entropy images.

10. Conclusion

In this article, we analyzed structural image representations for multi-modal registration. We formulated theoretical properties that such a structural representation should fulfill. Generally, the registration with such representations has the advantage that mono-modal similarity measures can be used for multi-modal registration. Further, we show that a more efficient optimization scheme becomes applicable. Throughout the article we introduced two possible structural representations, the entropy and Laplacian images. The entropy images only fulfill certain requirements of a relaxed version of the theoretical properties, however, they are fast to compute and lead to good alignments, making them a very practical solution. The Laplacian images fulfill all the theoretical requirements, the preservation of locality even optimally. These superior theoretical properties also lead to better registration results, however, these advantages come at a significantly higher computational cost. Finally, we think that the application of structural representations for multi-modal registration is an interesting research direction, which presents an alternative to the usage of sophisticated similarity metrics.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.media.2011.03.001.

References

Andronache, A., von Siebenthal, M., Székely, G., Cattin, P., 2008. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. Medical Image Analysis 12, 3–15.

Arandjelovic, O., Cipolla, R., 2007. A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution. In: ICCV.

Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computatu=ion 15.

Benhimane, S., Malis, E., 2004. Real-time image-based tracking of planes using efficient second-order minimization. In: IEEE/RSJ, pp. 943–948.

Buzug, T., Weese, J., Fassnacht, C., Lorenz, C., 1997. Image registration: convex weighting functions for histogram-based similarity measures. Lecture Notes in Computer Science, 203–212.

Collignon, A., Vandermeulen, D., Suetens, P., Marchal, G., 1995. 3D multi-modality medical image registration using feature space clustering. In: Computer Vision, Virtual Reality and Robotics in Medicine.

Du, S., Zheng, N., Ying, S., Liu, J., 2010. Affine iterative closest point algorithm for point set registration. Pattern Recognition Letters 31, 791–799.

Gerber, S., Tasdizen, T., Joshi, S., Whitaker, R., 2009. On the manifold structure of the space of brain images. In: MICCAI.

Haber, E., Modersitzki, J., 2007. Intensity gradient based registration and fusion of multi-modal images. Methods of Information in Medicine 46, 292–299.

Hamm, J., Davatzikos, C., Verma, R., 2009. Efficient large deformation registration via geodesics on a learned manifold of images. In: MICCAI, pp. 680–687.

Kadir, T., Brady, M., 2001. Saliency, scale and image description. International Journal of Computer Vision 45, 83–105.

Learned-Miller, E.G., 2006. Data driven image models through continuous joint alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 236–250.

Lee, C., Elgammal, A., 2007. Modeling view and posture manifolds for tracking. In: ICCV.

Lee, D., Hofmann, M., Steinke, F., Altun, Y., Cahill, N., Scholkopf, B., 2009. Learning similarity measure for multi-modal 3d image registration. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 186 –193.

Li, S., Wang, H., Soh, W., 1998. Robust estimation of rotation angles from image sequences using the annealing M-estimator. Journal of Mathematical Imaging and Vision 8, 181–192.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110.

Maintz, J., van den Elsen, P., Viergever, M., 1996. Comparison of edge-based and ridge-based registration of CT and MR brain images. Medical Image Analysis 1, 151–161.

Maintz, J., van der Elsen, P., Viergever, M., 1997. Registration of 3D medical images using simple morphological tools. In: International Conference on Information Processing in Medical Imaging, pp.204–217.

Mansoury, S., Pasha, E., 2008. Determination of maximum entropy probability distribution via Burg's measure of entropy. Applied Mathematical Sciences 2, 2851–2858.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1615–1630.

Modersitzki, J.,2009. FAIR: flexible algorithms for image registration. Society for Industrial and Applied Mathematics (SIAM).

Myronenko, A., Song, X., 2009. Image registration by minimization of residual complexity. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).

Padfield, D., 2010. Masked FFT registration. In: Conference on Computer Vision and Pattern Recognition (CVPR).

Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y., 2010. RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images. In: Conference on Computer Vision and Pattern Recognition (CVPR).

Penney, G., Weese, J., Little, J., Desmedt, P., Hill, D., Hawkes, D., 1998. A comparison of similarity measures for use in 2-d–3-d medical image registration. IEEE Transactions on Medical Imaging 17, 586–595.

Penney, G., Griffin, L., King, A., Hawkes, D., 2008. A novel framework for multi-modal intensity-based similarity measures based on internal similarity. SPIE, 6914.

Pizarro, D., Bartoli, A., 2007. Shadow resistant direct image registration. In: SCIA, pp. 928–937. .

Reddy, B., Chatterji, B., 1996. An FFT-based technique for translation, rotation, and scale-invariant image registration. IEEE Transactions on Image Processing 5, 1266–1271.

Roche, A., Malandain, G., Pennec, X., Ayache, N., 1998. The correlation ratio as a new similarity measure for multimodal image registration. In: Proc. of First Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98). Springer Verlag, Cambridge, USA, pp. 1115–1124.

Rohde, G.K., Wang, W., Peng, T., Murphy, R.F., 2008. Deformation-based nonlinear dimension reduction: applications to nuclear morphometry. In:ISBI.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326.

Rényi, A., 1961. On measures of entropy and information. In: Proc. Fourth Berkeley Symp. Math. Stat. and Probability, pp. 547–561.

Shechtman, E., Irani, M., 2007. Matching local self-similarities across images and videos. In: IEEE Conference on Computer Vision and Pattern Recognition.

Sizintsev, M., Derpanis, K., Hogue, A., 2008. Histogram-based search: a comparative study. In: IEEE Conference on Computer Vision and Pattern Recognition.

Tenenbaum, J., Silva, V., Langford, J., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319.

Wachinger, C., Navab, N., 2009. Similarity metrics and efficient optimization for simultaneous registration. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).

Wachinger, C., Navab, N., 2010a. Manifold learning for multi-modal image registration. In: 11st British Machine Vision Conference (BMVC).

Wachinger, C., Navab, N., 2010b. Structural image representation for image registration. In: CVPR Workshops, IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), San Francisco, California.

Wachinger, C., Wein, W., Navab, N., 2007. Three-dimensional ultrasound mosaicing. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Brisbane, Australia.

Wachinger, C., Mateus, D., Keil, A., Navab, N., 2010a. Manifold learning for patient position detection in MRI. In: IEEE International Symposium on Biomedical Imaging, Rotterdam, The Netherlands.

Wachinger, C., Yigitsoy, M., Navab, N., 2010b. Manifold learning for image-based breathing gating with application to 4D ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Beijing, China.

Wei, Y., Tao, L., 2010. Efficient histogram-based sliding window. In: IEEE Conference on Computer Vision and Pattern Recognition.

Wein, W., Brunke, S., Khamene, A., Callstrom, M., Navab, N., 2008. Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis 12, 577–585.

Wells, W., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multi-modal volume registration by maximization of mutual information. Medical Image Analysis 1, 35–51.

Yi, Z., Soatto, S., 2009. Nonrigid registration combining global and local statistics. In: CVPR, pp.2200–2207.

Zhang, Q., Souvenir, R., Pless, R., 2006. On manifold structure of cardiac MRI data: application to segmentation. In: CVPR, vol. 1, pp. 1092–1098.

Zöllei, L., Learned-Miller, E., Grimson, E., Wells, W., 2005. Efficient population registration of 3D data. In: Computer Vision for Biomedical Image Applications, ICCV.