

Multimodal Image Registration using Laplacian Commutators

Veronika A. Zimmer^{a,*}, Miguel Ángel González Ballester^{a,b}, Gemma Piella^a

^a*SIMBIOsys, Universitat Pompeu Fabra, Barcelona, Spain*

^b*ICREA, Barcelona, Spain*

Abstract

The fusion and combination of images from multiple modalities is important in many applications. Typically, this process consists of the alignment of the images and the combination of the complementary information. In this work, we focused on the former part and propose a multimodal image distance measure based on the commutativity of graph Laplacians. The eigenvectors of the image graph Laplacian, and thus the graph Laplacian itself, capture the intrinsic structure of the image's modality. Using Laplacian commutativity as a criterion of image structure preservation, we adapt the problem of finding the closest commuting operators to multimodal image registration. Hence, by using the relation between simultaneous diagonalization and commutativity of matrices, we compare multimodal image structures by means of the commutativity of their graph Laplacians. In this way, we avoid spectrum reordering schemes or additional manifold alignment steps which are necessary to ensure the comparability of eigenspaces across modalities. We show on synthetic and real datasets that this approach is applicable to dense rigid and non-rigid image registration. Results demonstrated that the proposed measure is able to deal with very challenging multimodal datasets and compares favorably to normalized mutual information, a de facto similarity measure for multimodal image registration.

Keywords: Image Registration, Simultaneous Diagonalization, Laplacian Commutator, Spectral Methods

1. Introduction

Image registration is an essential processing technique in applications that require the fusion or combination of information from different image sources and/or views [1, 2]. For example, in medical imaging, the diagnostic and treatment planning benefit often from the complementary information from multiple modalities, such as ultrasound and magnetic resonance [3] or from the study of temporal series of images for change detection of organs and tumors. The quality of the fused image depends highly on the accuracy of the images alignment [4], which is done using image registration. Image registration arises whenever images acquired from different sensors or from different scenes or at different times need to be compared

*Corresponding author

or combined for analysis. The objective of image registration is to find a plausible spatial transformation
 10 which aligns corresponding structures in two or more images. Typically, this is done by minimizing the
 distance between the images with respect to a transformation. Defining and measuring image distance or
 similarity is a difficult and highly application-dependent task, especially for multimodal images. Existing
 measures compare local intensities and features to find corresponding points in the images. When images
 are acquired with different acquisition parameters or systems, local intensities and intensity gradients can
 15 vary largely between the modalities. However, the global structure of such multimodal images is similar.

In this work, we propose to compare the structure of multimodal images by analyzing the commutativity
 of their graph Laplacians. We define the unnormalized graph Laplacian as $\mathbf{L} := \mathbf{D} - \mathbf{W}$ with matrices
 $\mathbf{W} = (w_{ij})$ and $\mathbf{D} = \text{diag}(\sum_{l=1}^n w_{il})$, where w_{ij} represents the similarity between image voxel or patch i
 and j . The eigenvectors of the image graph Laplacian, and thus the graph Laplacian itself, capture the
 20 intrinsic structure of the image, which is to some degree shared across modalities. Our work is based on the
 problem of finding the closest commuting operator [5] for manifold alignment. Relying on graph Laplacians
 as image structure descriptors and using Laplacian commutativity as a criterion for image structure preser-
 vation, we adapt the problem of finding the closest commuting operators to multimodal image registration.
 By indirectly comparing the eigenspaces of graph Laplacians from multimodal images through Laplacian
 25 commutativity, we avoid spectrum reordering schemes or additional steps of manifold alignment to ensure
 the comparability of eigenspaces across modalities.

When using spectral techniques in image registration, i.e., methods which use spectral decompositions
 to study the structure of data sets, the underlying assumption is that two images are similar if their corre-
 sponding graph Laplacians, as structural descriptors of the underlying manifold, have similar eigenvectors.
 30 However, the graph Laplacians computed independently from different images may not share the same
 eigenspace [6–8]. For monomodal data, and under the assumption that the underlying manifolds are iso-
 metric, the works in [6, 7] proposed spectrum reordering schemes. For multimodal data, where the isometry
 assumption is not valid anymore, the eigenspaces may greatly differ [9]. To deal with this, an additional
 step of manifold alignment is performed in [10] to ensure that the spaces are comparable across manifolds.
 35 Similarly, in [8] an eigenfunction change of basis [11] was done to transform the new coordinates of the
 structural representation of one image into the same coordinate space of the other.

An alternative way is to use simultaneous diagonalization. A set of matrices is simultaneously diagonal-
 izable, if there exists a single unitary matrix \mathbf{V} such that $\mathbf{V}^T \mathbf{A} \mathbf{V}$ is a diagonal matrix for each matrix \mathbf{A} in
 the set. This approach has been applied by [9, 12] to find a common eigenspace of multiple Laplacians. With
 40 that, they extended methods based on spectral geometry, like Laplacian eigenmaps and diffusion maps, to
 the multimodal setting.

In [5], the authors proposed a new paradigm for multimodal spectral geometry. Their approach is based
 on the fact that matrices are simultaneously diagonalizable if and only if they commute. Thus, instead of

finding a common eigenspace via simultaneous diagonalization, they searched for new graph Laplacians that
45 commute and are as close as possible to the original graph Laplacians. The eigendecomposition of those
closest commuting operators were then used for multimodal spectral analysis.

In this work, we propose to quantify structure resemblance between two multimodal images through
Laplacian commutators. If two multimodal images of the same object or scene are aligned, their structure
should be similar and hence also the Laplacian eigenspaces of their corresponding image graphs. Therefore,
50 we look for a (rigid or non-rigid) transformation that transforms one of the images such that the graph
Laplacians approximately commute. This implies that they are approximately diagonalizable and hence
their structure is similar. This is in line with the works [13, 14], where a similar approach was used
for image color transformations. To the best of our knowledge, this is the first work using simultaneous
diagonalization in the context of dense image registration.

55 The remainder of this paper is structured as follows: We review the most important works related to our
approach in Section 2. We introduce spectral methods in Section 3, including basic notions on manifolds
and graph theory. In Section 4, we overview the problem of approximate simultaneous diagonalization and
how this is related to the search for the closest commuting operators. In Section 5, we propose to use the
closest commuting operator to measure structure similarity in images and explain how to apply it to image
60 registration. Section 6 presents experimental results on synthetic and real datasets, followed by further
discussion and conclusions in Section 7.

2. Related work

Classical registration methods can be divided into feature- and intensity-based methods [15, 16]. In
feature-based registration techniques, image features, such as points and contours, are extracted and aligned,
65 and the distance measure is defined on the extracted features. While being often more efficient, such meth-
ods suffer from significant intensity variations in multimodal images, which makes it difficult to find enough
corresponding features for accurate matching. In intensity-based techniques, the distance measure is defined
directly on the intensity values of the image. A popular intensity-based measure for multimodal image reg-
istration is mutual information (MI) and its normalized version, normalized mutual information (NMI) [17].
70 Both estimate the shared information between the images to be registered by constructing intensity his-
tograms over the images. MI and NMI have been successfully applied in various applications. However,
MI-based measures still may fail in the presence of high intensity variations and intensity gradients [18].
Furthermore, they are non-convex measures and may have local maxima [19].

In recent years, structural image representations have gained interest for multimodal registration, since
75 they allow transforming images into a common space where modalities can be directly compared. They rely
on the assumption that internal self-similarities are preserved across modalities. The images are transformed

to new representations that capture such self-similarities and where efficient monomodal measures, such as the L1 or L2 distance, can be used for registration. Existing approaches have applied local gradient orientation [20, 21], local entropy [10] and local patch distance [22–24] to find correspondences between images.

80 However, in these approaches only the information of local neighborhoods is compared and, moreover, the local self-similarity pattern can vary significantly between corresponding points in images [18].

Manifold learning techniques, which are commonly used for population analysis [25], have also begun to be employed for image registration. Applied to individual images, such techniques allow obtaining a new image representation that can help in finding image features and correspondences. For example, PCA [26] 85 and Isomap [27] were employed in [28, 29], respectively, to extract features of monomodal images to be used for affine registration.

Spectral methods are specific manifold learning techniques, which use spectral decomposition to study the structure of data sets. Standard spectral methods find the eigenvectors of the Laplacian matrix of a graph, which is constructed on the original high-dimensional data space. A (suitable weighted) graph Laplacian can 90 be viewed as a discretized version of the Laplace-Beltrami operator on Riemannian manifolds [30, 31], whose eigenfunctions are known to characterize the structure of the manifold and can be therefore interpreted as a structural descriptor of the underlying manifold. Among other applications, spectral methods have been used for dimensionality reduction, clustering [32], and shape analysis and alignment [7, 33–36].

For image registration, spectral methods have received limited attention. In [8, 10], multimodal images 95 were represented by the first embedding coordinate obtained with Laplacian eigenmaps [37] and diffusion maps [38], respectively, which allowed the use of monomodal distance measures in the context of rigid [10] and non-rigid [8] image registration. In [6], the Log-Demons [39] was extended to deal with large deformations by incorporating geometric information through the comparison of the eigenvectors of the graph Laplacians constructed from images. However, this method was limited to monomodal images. The spectral 100 decomposition of a joint graph of two images or shapes was used in [18] for image matching and in [40] for surface matching.

3. Theoretical background

Let \mathcal{M} be a compact, k -dimensional Riemannian manifold (i.e., a differentiable manifold whose tangent spaces are equipped with an inner product), which is embedded into a d -dimensional Euclidean space. 105 Typically, $k \ll d$.

The structure of the manifold can be studied through the eigenfunctions of the Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ on \mathcal{M} [33, 41, 42], which reduces to the common Laplace operator $\Delta f = \sum_{i=1}^d \frac{\partial^2 f}{\partial r_i^2}$, $f : \mathcal{M} \rightarrow \mathbb{R}$, f twice differentiable and r_i being the Euclidean coordinates, if the manifold is an open subset of the Euclidean space. The eigenfunctions $\{v_i\}_{i=1}^{\infty}$ of $\Delta_{\mathcal{M}}$, satisfying the Laplacian eigenvalue problem $\Delta_{\mathcal{M}} v_i = \lambda_i v_i$,

110 form a complete and orthonormal basis of the Hilbert space $L^2(\mathcal{M})$ with respective eigenvalues $\{\lambda_i\}_{i=1}^\infty$,
 $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \infty$. Any $f \in L^2(\mathcal{M})$ has a Fourier expansion in the basis $\{v_i\}_{i=1}^\infty$, thus allowing Fourier
analysis on manifolds. The eigenvalues $\{\lambda_i\}_{i=1}^\infty$ are analogous to frequencies in the Euclidean case and the
eigenfunctions to the basis functions sine and cosine [33]. Low-frequency eigenfunctions, corresponding to
the smallest eigenvalues, describe the global structure of the manifold, while high-frequency eigenfunctions
115 capture the details [41, 42].

A common way to discretize a manifold is by a graph structure. An undirected weighted graph is an
ordered pair $G = (V, E)$ with a set of nodes $V = \{x_i\}_{i=1}^n$, edges $E \subseteq \{1, \dots, n\}^2$ and edge weights $w_{ij} \in \mathbb{R}_0^+$.
Such a graph is constructed from a data set by assigning each data point sampled on an underlying manifold
(possibly together with additional features) to a node. For instance, for analysis of large populations of mas-
120 sive images, the nodes consist of the images and a single graph for the image population is constructed [43].
For shape matching, the nodes are the points of the mesh, while for image registration, they consist of the
voxels or patches of the images and one graph is constructed per shape/image. The edge weights represent
a notion of similarity between data points. Two nodes x_i, x_j are connected with an edge if $w_{ij} \geq \theta$, $\theta \in \mathbb{R}_0^+$.
A popular choice is to use a Gaussian kernel, $w_{ij} = w(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$, where $\sigma > 0$ is
125 the kernel bandwidth [37]. However, other options are possible (cf. [44]).

As a discretization of the Laplace-Beltrami operator, the eigenvectors of a suitable graph Laplacian
discretize the eigenfunctions of the Laplace-Beltrami operator [31, 37, 45]. We define the unnormalized
graph Laplacian as $\mathbf{L}_u := \mathbf{D} - \mathbf{W}$, and the random-walk graph Laplacian as $\mathbf{L}_{rw} := \mathbf{D}^{-1}\mathbf{L}_u$ [30]. The
matrix $\mathbf{W} = (w_{ij})$ is called the adjacency matrix, and $\mathbf{D} = \text{diag}(\sum_{l=1}^n w_{il})$ the diagonal degree matrix,
130 containing the degree of each node, i.e., the sum of weights connected to the node. For the case of data
(e.g., voxels of the images) drawn from a non-uniform distribution, only the random-walk graph Laplacian
converges to the Laplace-Beltrami operator [30].

Both Laplacians are positive semi-definite matrices. They can be decomposed as $\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$,
where \mathbf{V} is the matrix of column eigenvectors of \mathbf{L} and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ is the diagonal matrix of
135 corresponding eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ for both $\mathbf{L} = \mathbf{L}_u$ and $\mathbf{L} = \mathbf{L}_{rw}$ ¹ [5, 9, 45].

A popular spectral embedding technique is the Laplacian eigenmap algorithm [37]. It captures the
intrinsic low-dimensional structure of a manifold by finding an optimal embedding which preserves local
neighborhoods. This can be posed as the minimization problem

$$\min_{\mathbf{Y}^T \mathbf{D} \mathbf{Y} = \mathbf{I}} \sum_{ij} \|y_i - y_j\|^2 w_{ij},$$

where $\mathbf{Y} = [y_1, \dots, y_n] \in \mathbb{R}^{d \times n}$ are the embedding coordinates and w_{ij} the similarity between input data
points, thus the edge weight between nodes x_i and x_j . By using the above objective function, neighboring

¹Note that their eigenvectors and -values are not the same: $\mathbf{V}_u \neq \mathbf{V}_{rw}$, though they are related: the eigenvectors of $\mathbf{L} = \mathbf{L}_{rw}$
are generalized eigenvectors of $\mathbf{L} = \mathbf{L}_u$, i.e., $\mathbf{L}_u \mathbf{V}_{rw} = \lambda \mathbf{D} \mathbf{V}_{rw}$.

points in the original space (w_{ij} large) receive a penalty if they are mapped far apart in the embedding space ($y_i - y_j$ large). Thus, by minimization of the objective function local neighborhoods are preserved. It can be shown [37] that solving this minimization problem is equivalent to the generalized eigenvalue problem of the graph Laplacian, $\mathbf{L}\mathbf{v} = \lambda\mathbf{D}\mathbf{v}$, whose eigenvectors \mathbf{v}_i give the optimal embedding, with a diagonal degree matrix \mathbf{D} .

4. Simultaneous diagonalization of graph Laplacians

In various applications, such as shape matching or multimodal data analysis, the structure of two or more manifolds have to be compared. Using spectral methods, this reduces to embedding the geometrical objects into the eigenspace of the Laplace-Beltrami operator and comparing those new representations. If isometric objects are considered, the objects share the same spectral coordinates, since the Laplace-Beltrami eigenspace is invariant under isometric transformations [46]. Isometric objects are, e.g., two shapes representing the same object, but in different poses. However, even if the graph Laplacians from isometric shapes share the same eigenbasis, they may not be directly comparable. This is due to three reasons. First, the eigenvectors are only defined up to a sign: if \mathbf{v} is an eigenvector of \mathbf{L} , $\mathbf{L}\mathbf{v} = \lambda\mathbf{v}$, then $\mathbf{L}(-\mathbf{v}) = \lambda(-\mathbf{v})$, i.e., $-\mathbf{v}$ is an eigenvector with the same eigenvalue. Second, if the multiplicity of an eigenvalue is greater than one, the eigenvectors corresponding to this eigenvalue may be in a different order or subject to an orthonormal transformation. Spectrum reordering schemes have been proposed to deal with this type of eigenspace incompatibility [6, 7, 46, 47]. Third, since many Laplacian discretizations are highly mesh sensitive, if the sampling sizes are different in the isometric manifolds, the eigenspaces may not be comparable anymore. On the other hand, in the case of non-isometric manifolds, the eigenvectors of the individual graph Laplacians differ strongly and are therefore inconsistent with each other. In particular, data sets from different modalities lie on non-isometric manifolds, which makes it challenging to use spectral methods in the multimodal case.

As commented in Section 1, one possible approach to overcome this problem is a change of eigenbasis [6, 8]. Another approach is to perform simultaneous diagonalization of the graph Laplacians so that they share the same set of eigenvectors. Two real $n \times n$ matrices $\mathbf{A}_1, \mathbf{A}_2$ are said to be *simultaneously diagonalizable* if there exists a unitary matrix \mathbf{V} (i.e., $\mathbf{V}^T\mathbf{V} = \mathbf{I}$) such that both $\mathbf{V}^T\mathbf{A}_1\mathbf{V}$ and $\mathbf{V}^T\mathbf{A}_2\mathbf{V}$ are diagonal matrices. Thus, \mathbf{V} is the joint eigenbasis of \mathbf{A}_1 and \mathbf{A}_2 . If the matrices $\mathbf{A}_1, \mathbf{A}_2$ are not simultaneously diagonalizable, which is the case in many applications due to noise, discretization and other circumstances, one seeks to find approximately joint eigenspaces.

Several numerical methods for the approximate simultaneous diagonalization of a series of square symmetric matrices have been proposed [48–53]. They search for a matrix $\mathbf{V} \in \mathbb{R}^{n \times n}$ such that $\mathbf{V}^T\mathbf{A}_1\mathbf{V}$ and $\mathbf{V}^T\mathbf{A}_2\mathbf{V}$ are as close to diagonal matrices as possible. This can be done by solving the optimization problem

$$J(\mathbf{A}_1, \mathbf{A}_2) = \min_{\substack{\mathbf{V} \in \mathbb{R}^{n \times n} \\ \mathbf{V}^T \mathbf{V} = \mathbf{I}}} \sum_{i=1}^2 \|\mathbf{V}^T \mathbf{A}_i \mathbf{V} - \text{Diag}(\mathbf{V}^T \mathbf{A}_i \mathbf{V})\|_F^2, \quad (1)$$

where $\text{Diag}(A)$ denotes a matrix obtained by setting the off-diagonal elements of A to zero (following the notation in [5]). The matrix norm $\|\mathbf{A}\|_F = \left(\sum_{ij} |a_{ij}|^2\right)^{1/2}$ is the *Frobenius-norm*. The existing numerical methods differ in the way to solve the optimization problem and in the restrictions which are set to matrix \mathbf{V} .

These joint eigenspaces can be used to analyze simultaneously the structure of multimodal datasets. This was done in [9, 12, 54], where the eigenvectors and eigenvalues of a single graph Laplacian were replaced by the joint eigenvectors and eigenvalues of multiple graph Laplacians.

Simultaneous diagonalization of matrices is closely related to commutativity of matrices. Two matrices $\mathbf{A}_1, \mathbf{A}_2$ are simultaneously diagonalizable by a unitary similarity transformation if and only if they commute [55], e.g., $[\mathbf{A}_1, \mathbf{A}_2] := \mathbf{A}_1 \mathbf{A}_2 - \mathbf{A}_2 \mathbf{A}_1 = \mathbf{0}$. This can be extended to the approximate case. In [56], the authors showed that for symmetric matrices $\mathbf{A}_i \in \mathbb{R}^{n \times n}$ with $\|\mathbf{A}_i\|_F = 1, i = 1, 2$, there exist two functions $\delta_1(x), \delta_2(x)$ with $\lim_{x \rightarrow 0} \delta_i(x) = 0, i = 1, 2$, such that

$$\delta_1(\|\mathbf{A}_1, \mathbf{A}_2\|_F) \leq J(\mathbf{A}_1, \mathbf{A}_2) \leq n\delta_2(\|\mathbf{A}_1, \mathbf{A}_2\|_F).$$

In other words, almost commuting matrices (matrices that approximately commute, i.e., $[A, B] \leq \delta$ for some small $\delta > 0$ [57]) are almost simultaneously diagonalizable.

Together with Lin's theorem [57], stating that almost commuting matrices are close to commuting matrices, this implies that almost simultaneously diagonalizable matrices are close to simultaneously diagonalizable matrices.

Thus, instead of finding a matrix \mathbf{V} , which approximately diagonalizes both matrices \mathbf{A}_1 and \mathbf{A}_2 as in Eq. (1), one can seek commuting matrices $\tilde{\mathbf{A}}_1, \tilde{\mathbf{A}}_2$ that are as close as possible to $\mathbf{A}_1, \mathbf{A}_2$, respectively [5]. This problem can be formulated as

$$C(\mathbf{A}_1, \mathbf{A}_2) = \min_{\tilde{\mathbf{A}}_1, \tilde{\mathbf{A}}_2} \sum_{i=1}^2 \|\tilde{\mathbf{A}}_i - \mathbf{A}_i\|_F^2 \quad \text{s.t.} \quad \|\tilde{\mathbf{A}}_1 \tilde{\mathbf{A}}_2 - \tilde{\mathbf{A}}_2 \tilde{\mathbf{A}}_1\|_F^2 = 0. \quad (2)$$

The solutions to problem (2), the matrices $\tilde{\mathbf{A}}_1, \tilde{\mathbf{A}}_2$, are commuting matrices and therefore simultaneously diagonalizable.

It has been shown [5] that the problems of Eqs. (1) and (2) are equivalent in the sense that both problems have the same minimal value ($J(\mathbf{A}_1, \mathbf{A}_2) = C(\mathbf{A}_1, \mathbf{A}_2)$), the minimizers of $J(\mathbf{A}_1, \mathbf{A}_2)$ are the joint eigenvectors of the minimizers of $C(\mathbf{A}_1, \mathbf{A}_2)$, and $\tilde{\mathbf{A}}_i, i = 1, 2$, are the closest commuting matrices of the approximate joint eigenvectors $\hat{\mathbf{V}}$ of \mathbf{A}_i with $\tilde{\mathbf{A}}_i = \hat{\mathbf{V}} \text{Diag}(\hat{\mathbf{V}}^T \mathbf{A}_i \hat{\mathbf{V}}) \hat{\mathbf{V}}^T$.

A key difference between Eqs. (1) and (2) is that the former optimizes with respect to the joint eigenbasis, while the latter does it with respect to the commuting matrices, thus having an explicit control over the

structure of the matrices. This is an advantage when dealing with graph Laplacians since it allows restricting the closest commuting matrices to plausible graph Laplacians. In contrast, in simultaneous diagonalization, if $\hat{\mathbf{V}}$ is the joint eigenbasis that approximately diagonalizes two graph Laplacians $\mathbf{L}_1, \mathbf{L}_2$, then the almost commuting graph Laplacians $\hat{\mathbf{L}}_i = \hat{\mathbf{V}}\text{Diag}(\hat{\mathbf{V}}^T \mathbf{L}_i \hat{\mathbf{V}}) \hat{\mathbf{V}}^T \approx \mathbf{L}_i$, $i = 1, 2$, do not necessarily fulfill the properties of graph Laplacians (e.g., being symmetric and positive semi-definite). This was pointed out in [5], where they obtained the closest commuting matrices among the subset of graph Laplacians by reformulating (2) as:

$$\min_{\{\mathbf{u}_k \geq 0\}_{k=1}^2} \left(\sum_{k=1}^2 \|\tilde{\mathbf{L}}_k(\mathbf{u}_k) - \mathbf{L}_k\|_F^2 + \gamma \|\tilde{\mathbf{L}}_1(\mathbf{u}_1)\tilde{\mathbf{L}}_2(\mathbf{u}_2) - \tilde{\mathbf{L}}_2(\mathbf{u}_2)\tilde{\mathbf{L}}_1(\mathbf{u}_1)\|_F^2 \right), \quad (3)$$

where $\gamma \in \mathbb{R}^+$ is a weighting parameter to determine the influence of the structure preserving term (first term) over the commuting term (second term). Here, $\mathbf{L}_1, \mathbf{L}_2$ are the graph Laplacians and $\tilde{\mathbf{L}}_1(\mathbf{u}_1), \tilde{\mathbf{L}}_2(\mathbf{u}_2)$ are the almost commuting graph Laplacians, parametrized by the weights $\mathbf{u}_k = (u_1^k, \dots, u_{M_k}^k)$, $k = 1, 2$, of the corresponding adjacency matrices $\tilde{\mathbf{W}}_k$, with $M_k = \frac{1}{2}|\tilde{E}_k|$ (due to symmetry) and

$$\tilde{w}_{ij}^k(\mathbf{u}_k) = \begin{cases} u_i^k & \text{if } (i_l, j_l) \text{ or } (j_l, i_l) \in \tilde{E}_k \\ 0 & \text{otherwise.} \end{cases}$$

The new Laplacians $\tilde{\mathbf{L}}_1, \tilde{\mathbf{L}}_2$ are called the *closest commuting operators* (CCO).

This approach to multimodal spectral geometry has been successfully applied to shape matching and clustering [5].

5. Application to multimodal image registration

Let $I^F, I^M : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$ be the fixed and moving image, respectively, and $\phi_{\mathbf{p}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ a transformation with parameters \mathbf{p} , which maps points from the moving image space to the fixed image space. Image registration finds a $\phi_{\mathbf{p}}$ such that $I^F \approx I^M \circ \phi_{\mathbf{p}}$. The registration problem can be formulated as an optimization problem, where the chosen distance or dissimilarity measure is minimized with respect to a transformation. In general, the problem is ill-posed and a regularizer is added to the cost function of the optimization to provide some control over the transformation, e.g., regarding differentiability properties. The cost function to be optimized during registration is therefore

$$\mathcal{J}(\phi_{\mathbf{p}}) = \mathcal{D}(I^F, I^M; \phi_{\mathbf{p}}) + \mu \mathcal{R}(\phi_{\mathbf{p}}), \quad (4)$$

with a distance measure \mathcal{D} , a regularizer \mathcal{R} and a regularization penalty weight $\mu \in \mathbb{R}$. To solve the registration problem, Eq. (4) is optimized in an iterative manner.

We propose to use a (non-metric) distance measure based on the image structure similarity, conveyed by the similarity of the eigenspaces of the image graph Laplacians.

In this work, we follow the interpretation of [58] that an image is the embedding map of a higher dimensional manifold. For each image $I : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$, we construct the corresponding image graph $G = (V, E)$ with a set of nodes $V = \{x_i\}_{i=1}^n$ and edges E . Each node x_i consists of a voxel $u_i \in \Omega$ and a vector $\mathbf{I}(u_i) = (I(u_i^1), \dots, I(u_i^N))$, $\mathbf{I}(u_i) \in \mathbb{R}^N$, of image intensities within a patch centered at u_i , where $u_i^j \in \Omega$ are the voxels in the patch. If $N = 1$, the node consists only of voxel u_i and intensity $I(u_i)$, and the resulting graph is called voxel-based graph. For $N > 1$, the graph is patch-based.

We compute the weights by using (patch-based) bilateral filtering [59]:

$$w_{ij} = w(x_i, x_j) = \exp \left(- \left[\frac{\|\mathbf{I}(u_i) - \mathbf{I}(u_j)\|^2}{2\sigma_1^2} + \beta \frac{\|u_i - u_j\|^2}{2\sigma_2^2} \right] \right), \quad (5)$$

so that voxels with similar intensities have a large weight only if they lie close in space. The parameters σ_1 and σ_2 play the role of scale parameters. Interpreted within the scale-space theory presented in [60], σ_1 corresponds to the intensity scale or resolution in the domain of the sampled value (i.e., tonal scale) and σ_2 to the spatial scale (i.e., inner scale) at which the image is observed. The parameter $\beta \in \mathbb{R}_0^+$ corrects for scale differences between the intensities and spatial coordinates by balancing the influence of the spatial component. The corresponding image graph Laplacian (unnormalized and random-walk) is then constructed as explained in Section 3.

For image registration (see Eq. (4)), we look for a transformation $\phi_{\mathbf{p}}$ that deforms the moving image I^M such that it is most similar to a fixed image I^F . We assume that the images have the same size and origin, such that they share the same voxel coordinates. From I^F and $I^M \circ \phi_{\mathbf{p}}$, we obtain their corresponding graph Laplacians \mathbf{L}^F and \mathbf{L}^M , respectively, as detailed in Section 3. Since the moving image changes with the transformation parameters \mathbf{p} , we have that $w_{ij}^M = w_{ij}^M(\mathbf{p})$ and $\mathbf{L}^M = \mathbf{L}^M(w_{ij}^M(\mathbf{p}))$. For simplicity of notation, we suppress the dependence of the weights and write $\mathbf{L}^M(\mathbf{p})$.

As presented in Section 3, the structure of a Riemannian manifold can be characterized by the eigenfunctions of the Laplace-Beltrami operator, which can be discretized through the graph Laplacian. Therefore, one can interpret the graph Laplacian of an image as a structure descriptor of the image and the similarity of graph Laplacian eigenspaces as a criterion of structural similarity between two images. By enforcing the graph Laplacians of a fixed and moving image during registration to have a joint eigenspace, the structure of the images can be aligned. Note that the joint eigenspace stands here for a basis which simultaneously diagonalizes the graph Laplacians of both the fixed and moving image, in contrast with an eigenspace of the Laplacian of a joint graph constructed using both images (see, e.g., [18]).

One approach to register one image to another would be to search for a transformation such that the graph Laplacians of both images are simultaneously diagonalizable, similar to the ideas in [9, 12]. To do this, a possible image distance function could be

$$\mathcal{D}(\mathbf{p}) = \mathcal{D}(I^F, I^M; \phi_{\mathbf{p}}) := \|\mathbf{V}^T \mathbf{L}^F \mathbf{V} - \text{Diag}(\mathbf{V}^T \mathbf{L}^F \mathbf{V})\|_F^2 + \|\mathbf{V}^T \mathbf{L}^M(\mathbf{p}) \mathbf{V} - \text{Diag}(\mathbf{V}^T \mathbf{L}^M(\mathbf{p}) \mathbf{V})\|_F^2. \quad (6)$$

Here, Eq. (1) has been used to approximate simultaneous diagonalization. For a minimizer $\phi_{\mathbf{p}}^*$, the matrix \mathbf{V} approximately diagonalizes both \mathbf{L}^F and $\mathbf{L}^M(\mathbf{p})$.

Figure 2 in Section 6.3 illustrates how the joint eigenspaces of aligned images diagonalize the image graph Laplacians, in contrast to joint eigenspaces of misaligned images. This indicates that finding a transformation such that the fixed and moving image graph Laplacians are almost simultaneously diagonalizable (by minimizing Eq. (6)) is a suitable criterion for image registration.

However, the simultaneous diagonalization of large matrices is computationally expensive and can take up to hours for large matrices. Typically, a registration algorithm is an iterative process, hence it is not practical to compute the joint eigenspaces at each iteration. In [9], a faster method for sparse Laplacians was presented. We overcome the high computational complexity by using the relation between simultaneous diagonalization and commuting matrices described in Section 4.

Our approach is to find the transformation parameters \mathbf{p} , such that the two image graph Laplacians \mathbf{L}^F and $\mathbf{L}^M(\mathbf{p})$ almost commute, which has been shown to be equivalent to approximate simultaneous diagonalization [56]. Additional regularization is used to constrain the matrix to preserve the structure of the fixed graph Laplacian.

Thus, we define the *closest commuting operator distance* (CCOD) as

$$\mathcal{D}(\mathbf{p}) = \mathcal{D}(I^F, I^M; \phi_{\mathbf{p}}) := \alpha \|\mathbf{L}^F \mathbf{L}^M(\mathbf{p}) - \mathbf{L}^M(\mathbf{p}) \mathbf{L}^F\|_F^2 + (1 - \alpha) \|\mathbf{L}^F - \mathbf{L}^M(\mathbf{p})\|_F^2. \quad (7)$$

The first term $\mathcal{D}_1(\mathbf{p}) := \|\mathbf{L}^F \mathbf{L}^M(\mathbf{p}) - \mathbf{L}^M(\mathbf{p}) \mathbf{L}^F\|_F^2$ measures how well the Laplacians commute. The second term $\mathcal{D}_2(\mathbf{p}) := \|\mathbf{L}^F - \mathbf{L}^M(\mathbf{p})\|_F^2$ ensures that the moving image has a Laplacian similar to the fixed image, thus enforcing that the resulting $\mathbf{L}^M(\mathbf{p})$ preserves the fixed image structure. The weighting parameter $\alpha \in \mathbb{R}^+$ controls the influence of both terms to the overall distance. This problem differs from the closest commuting operator problem in Eq. (3), used in [5], in three aspects. First, the main difference lies in the interpretation of the underlying problem. In [5], the CCO of Laplacians of two modalities were computed to obtain the eigenvectors of the CCO, the joint eigenvectors, as a more meaningful representation of the data in comparison to the individual eigenvectors of the Laplacians. Due to differences in modalities and presence of noise, the Laplacians are not exactly simultaneously diagonalizable and thus an approximate solution is sought. Thus, both Laplacians are modified (to find this approximation) but not the actual data. In our formulation, we not only want to obtain the CCO's of two modalities, but, in addition, one modality is geometrically transformed to enforce the commutativity (as a proxy for structure similarity). This is done by parametrizing the Laplacian of the moving image \mathbf{L}^M with the graph weights and the transformation parameters.

Second, in contrast to Eq. (3), in Eq. (7) only the moving image graph Laplacian \mathbf{L}^M is modified and \mathbf{L}^F is fixed. Instead of modifying both Laplacians slightly, such that they commute, we modify only \mathbf{L}^M to transform the underlying image data by a spatial transformation $\phi_{\mathbf{p}}$ such that \mathbf{L}^F and $\mathbf{L}^M(\mathbf{p})$ commute.

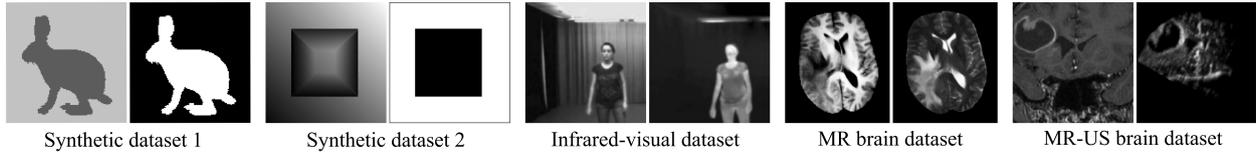


Figure 1: Examples of image pairs from the different datasets used for image registration.

Third, in Eq. (3), the CCOs are compared to the original Laplacians such that they are as close as possible to them. For our application, we do not want $\mathbf{L}^M(\mathbf{p})$ as close as possible to \mathbf{L}^M . The moving image is transformed during the registration process and in the end, $\mathbf{L}^M(\mathbf{p})$ might be very different to \mathbf{L}^M . Therefore, $\mathbf{L}^M(\mathbf{p})$ is compared to \mathbf{L}^F in the structure preservation term of Eq. (7) to ensure that $\mathbf{L}^M(\mathbf{p})$ has approximately the structure of the Laplacian of the fixed image.

We thus use CCOD (Eq. (7)) as a (non-metric) distance measure for image registration. Note that CCOD is asymptotically a (pseudo-)metric (cf. Appendix A).

We solve $\min_{\mathbf{p}} \mathcal{D}(\mathbf{p})$ using gradient-descent optimization methods. The computation of the derivatives of \mathcal{D} with respect to the transformation parameters \mathbf{p} for both normalized and unnormalized graph Laplacians is detailed in Appendix B.

6. Experimental results

We evaluated CCOD on both synthetic and real multimodal datasets. We especially chose the datasets to evaluate the performance of CCOD in challenging contexts (e.g. images with different intensity distribution, large deformations and missing structures).

We carried out experiments comparing the eigenspaces of images as well as distance curves and registration accuracy for global and local transformations (rigid and B-Spline). We compared CCOD to normalized mutual information (NMI) and local NMI (LNMI) [61], which are state-of-the-art distance measures for multimodal image registration.

For the registration experiments using transformations modeled with B-Splines, we used a regularizer penalizing second order derivatives:

$$\mathcal{R}(\phi) = \|\nabla_{\mathbf{x}}\phi\| = \sum_{i,j} \left(\frac{\partial^2 \phi(\mathbf{x})}{\partial x_i \partial x_j} \right)^2, \text{ with } \phi : \mathbb{R}^d \rightarrow \mathbb{R}^d \text{ being a spatial transformation (see Section 5).}$$

6.1. Parameters

The parameters of CCOD were chosen semi-automatically. The registration experiments were performed using a multiresolution strategy with three levels $i = 1, 2, 3$, and corresponding image sizes of 2^{4+i} in all dimensions. For all experiments, we constructed patch-based image graphs using a cubic patch with edge

length 2^i for image level i . We estimated σ_1 and σ_2 in the Gaussian kernel of Eq. (5) by:

$$\begin{aligned}\sigma_1 &= \underset{\substack{i=1,\dots,n \\ j=N_{\mathbf{I}(u_i)}(k)}}{\text{median}} \|\mathbf{I}(u_i) - \mathbf{I}(u_j)\|^2, \\ \sigma_2 &= \underset{\substack{i=1,\dots,n \\ j=N_{u_i}(k)}}{\text{median}} \|u_i - u_j\|^2,\end{aligned}\tag{8}$$

where $N_{\mathbf{I}(u_i)}(k)$, $N_{u_i}(k)$ are the indices of the k th nearest neighbor of vector $\mathbf{I}(u_i)$ and voxel u_i , respectively. We tested different values of k by taking $\nu\%$ of the number of nodes in the graph with $\nu \in \{5, 10, 50, 70, 100\}$. The choice of β in Eq. (5), unless otherwise stated, was computed using

$$\beta = \frac{\max_{i,j=1,\dots,n} \|\mathbf{I}(u_i) - \mathbf{I}(u_j)\|^2}{\max_{i,j=1,\dots,n} \|u_i - u_j\|^2}.\tag{9}$$

275 We chose α in Eq. (7) as $\alpha = \frac{\mathcal{D}_2}{\gamma \cdot \mathcal{D}_1 + \mathcal{D}_2}$, where \mathcal{D}_1 is the first term (the commutativity term; weighted by α) and \mathcal{D}_2 the second term (the structure preservation term; weighted by $1 - \alpha$). When $\gamma = 1$, α accounts for the different ranges of \mathcal{D}_1 and \mathcal{D}_2 and both terms contribute equally to the final distance. For all experiments we chose $\gamma = 0.7$ (the commutativity term has a higher weight), which gave the best results.

280 We used the unnormalized graph Laplacian only for the synthetic dataset, where the multimodal images have approximately the same intensity distribution (as specified in Section 6.2.1). For datasets, in which the ranges of the images to register are very different, we used the normalized version.

The most sensitive parameter is ν , which gives the percentage of number of nodes for determining the number of neighbors k for estimating the scale parameters σ_1 and σ_2 . This was the only parameter that had to be adapted to each multimodal dataset. Empirically, we found that ν should be small, when both 285 multimodal images have a lot of texture. However, for the infrared-visual datasets $\nu = 100$ gave the best result. Therefore, this parameter has to be carefully tuned for registration, similar to the fine tuning of the number of histogram bins for NMI.

The algorithms were implemented in C++ using the Insight Segmentation and Registration Toolkit (ITK²) and ran on a quad-core Intel (2 GHz CPU, 8 GB RAM). The average computation time when using 290 the proposed distance measure for 2D and 3D rigid registration was about 40 seconds and 21 minutes, respectively, and for non-rigid registration 10 minutes and 280 minutes, respectively.

6.2. Datasets

For the evaluation of CCOD, we used two different synthetic datasets, a dataset of infrared and visible images of pedestrians and two clinical datasets. The selected datasets show a variety of multimodal images, 295 ranging from simple binary images (first synthetic dataset) to multimodal images with highly different

²www.itk.org

intensity distributions and missing structures (all real datasets, but in particular the infrared-visual and second clinical dataset), to best explore the properties and limitations of the proposed distance measure. Examples of image pairs from the different datasets are shown in Fig. 1.

6.2.1. Synthetic data

300 *Data.* The first synthetic dataset consists of seven multimodal binary images showing various shapes (original databases from [62, 63]). Each shape has two modalities: one with a black background and a white object, the other with a light grey background and a dark grey object. For the second dataset, we used six synthetic phantoms similar to those employed in [64]. Both the fixed and moving image contain a geometrical object (a circle and a square, respectively), which is black with a white background or with an
305 intensity gradient.

The aim of using this dataset is to illustrate the behavior of CCOD for image registration using simple toy examples (see Sections 6.3, 6.4 and 6.5.1). NMI is known to perform badly when the images do not have the same intensity distribution as in the second synthetic dataset. LNMI, however, was proposed to overcome those short-comings.

310 *Parameters.* For the first synthetic dataset, we used $\nu = 100$, meaning that k in Eq. (8) is equal to the number of patches or voxels in the graph, respectively. Since the images are binary, using the pairwise intensity differences was enough to build the image graph, thus we set $\beta = 0$ in Eq. (5). For the second synthetic dataset, best results were obtained with $\nu = 5$ to compute k for cases where both images have texture and $\nu = 50$ where the moving image is binary. The parameter β was computed using Eq. (9). For all
315 the synthetic experiments, results are shown for the unnormalized Laplacian (similar results were obtained for the normalized one).

6.2.2. Visual-infrared data

Data. We used a dataset obtained from [65] consisting of eight pairs of infrared-visual images of pedestrians. The images are single frames of stereo videos acquired with a FLIR Thermovision A40M (infrared videos)
320 and a Sony XCD-710CR Sensor (visual images). We extracted a region of interest around the pedestrians that includes background features in the visual images, that are not visible in the infrared images. Apart from the different intensity distributions, these missing structures are challenging for intensity-based image registration. By using this dataset, we assess how well CCOD can align global image structures when locally structures are missing in one of the images.

325 *Parameters.* For the visual-infrared dataset, parameters for the random-walk graph Laplacian were computed as indicated in Section 6.1, with $\nu = 100$ to compute k .

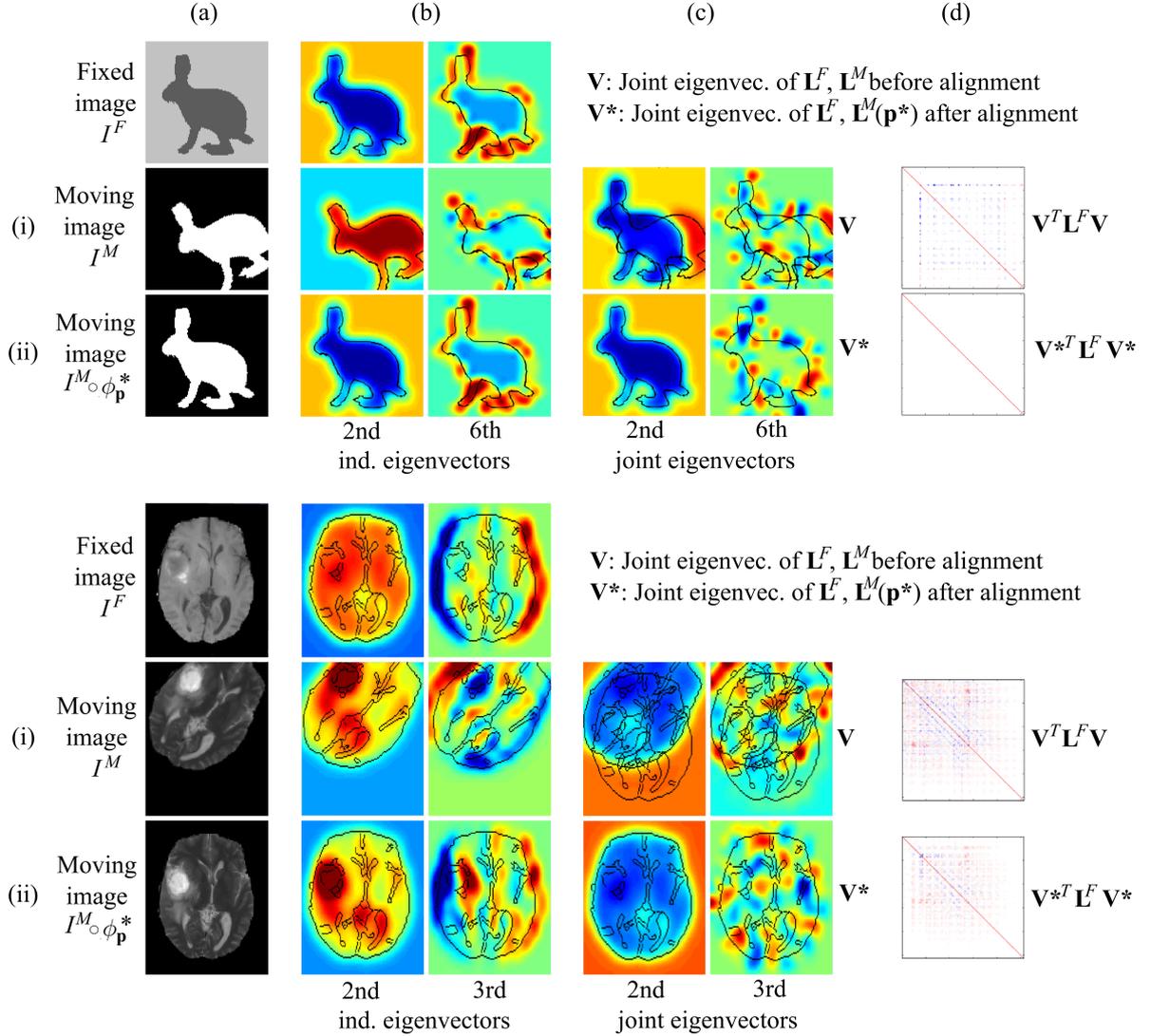


Figure 2: Individual and joint eigenvectors (with image contours) of (i) graph Laplacians $\mathbf{L}^F, \mathbf{L}^M$ of an image pair I^F, I^M before alignment and (ii) graph Laplacians $\mathbf{L}^F, \mathbf{L}^M(\mathbf{p}^*)$ of an image pair $I^F, I^M \circ \phi_p^*$ after alignment using two datasets. (a): fixed and moving images I^F, I^M before alignment and moving image after alignment ($I^M \circ \phi_p^*$, bottom row); (b): two individual eigenvectors of corresponding graph Laplacians; (c): joint eigenvectors \mathbf{V} of graph Laplacians $\mathbf{L}^F, \mathbf{L}^M$ before alignment (middle row) and \mathbf{V}^* of $\mathbf{L}^F, \mathbf{L}^M(\mathbf{p}^*)$ after alignment (bottom row); (d) matrices $\mathbf{V}^T \mathbf{L}^F \mathbf{V}$ and $\mathbf{V}^{*T} \mathbf{L}^F \mathbf{V}^*$. The matrix $\mathbf{V}^{*T} \mathbf{L}^F \mathbf{V}^*$ is a diagonal matrix, indicating that the corresponding graph Laplacians after alignment are simultaneously diagonalizable as opposed to the graph Laplacians before alignment.

6.2.3. Clinical data

Data. As a first clinical dataset (T1-T2 MR Brain Dataset), we used data obtained from the Brain Tumor segmentation challenge (MICCAI 2012 BRATS Challenge [66]). The dataset consists of T1 and T2 weighted magnetic resonance (MR) images of ten patients with ground truth alignment.

As a second clinical dataset (MR-US Brain Dataset), we used a public database of pre-operative T1-weighted MR and intra-operative (pre-resection) freehand ultrasound (US) images of ten tumor patients (BITE: Brain Images of Tumors for Evaluation [67]). Corresponding landmarks for each MR/US image pair selected independently by three experts are provided.

335 By using clinical data, the aim is to test the performance of CCOD on image data with rich texture (T1-T2) and on data containing the same object, but with many structural and textural differences (MR-US).

Parameters. For both clinical datasets, parameters for the random-walk graph Laplacian were computed as indicated in Section 6.1, with $\nu = 10$ to compute k .

6.3. Simultaneous diagonalization for image registration

340 We performed experiments to support the use of simultaneous diagonalization (e.g., through Eq. (6) and CCO (Eq. (7)) for pairwise image registration.

Figure 2 shows individual and joint eigenspaces before and after image alignment of an image pair of the first synthetic (top) and MR brain dataset (bottom). As we can see, the joint eigenspaces of the image graph Laplacians L^F and L^M (moving image before alignment) are very different and the eigenvectors \mathbf{V} do not diagonalize the graph Laplacians (middle row). After image alignment with a transformation $\phi_{\mathbf{p}}^*$ (bottom row), the joint eigenspaces are similar to the individual eigenspaces of L^F and $L^M(\mathbf{p}^*)$ and the matrix $\mathbf{V}^{*T} \mathbf{L}^F \mathbf{V}^*$ is closer to a diagonal matrix than before alignment. In particular, the distance in Eq. (6) is 9.73 before and 0.0 after alignment (synthetic image pair) and 33.52 before and 18.72 after alignment (MR brain image pair). Even for aligned images, Eq. (6) is not zero, meaning that the image graph Laplacians of I^F and $I^M \circ \phi_{\mathbf{p}}^*$ are not perfectly simultaneously diagonalizable. This can be explained by factors such as noise, but also by the different intensity distributions of the two image modalities.

345
350

6.4. CCOD for image registration

To assess the behavior of CCOD for measuring the similarity between images in pairwise registration, we compare its distance curves with those obtained by NMI and show its value as objective function in a registration experiment.

355

Figure 3 displays the distance curves obtained by horizontally translating an image over another, using an image pair of the first and second synthetic dataset. The distance values measured by both NMI and CCOD are plotted for each translation. For the first image pair (Fig. 3, left), the curves based on both NMI and CCOD have a global minimum³ for zero translation. However, while the distance curve of NMI has multiple local extrema, the curve obtained by CCOD is smooth with a single extremum at zero translation.

360

³NMI is a similarity measure. Maximizing NMI is equivalent to minimize the negative of NMI (which can be interpreted as a distance measure).

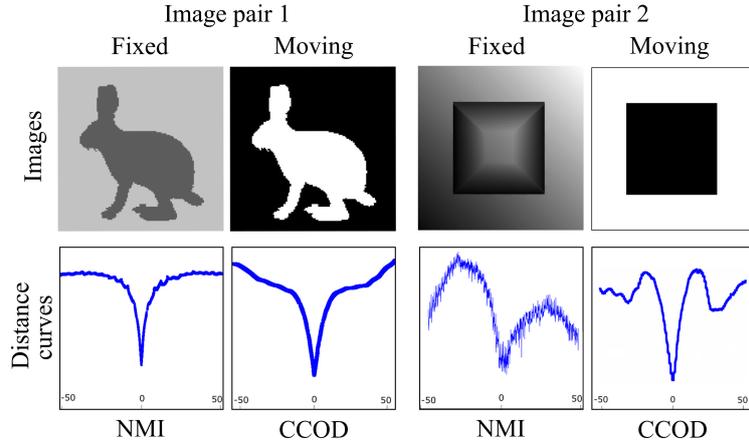


Figure 3: The moving images are translated horizontally over the fixed images for each image pair, respectively, to obtain distance curves using NMI and CCOD.

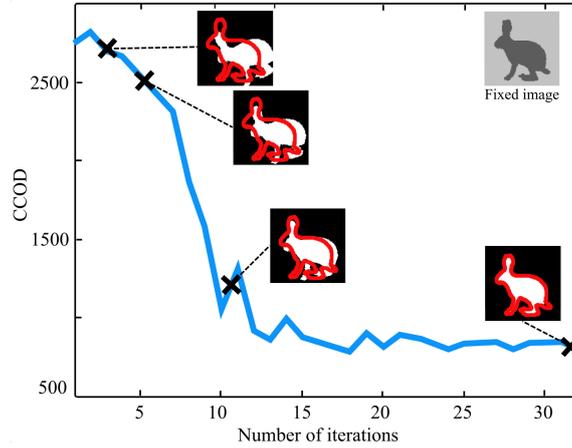


Figure 4: Registration example using CCOD on an image pair of the first synthetic dataset. The iterations of the optimization process for the registration of a moving to a fixed image are plotted against the value of CCOD.

For the second image pair (Fig. 3, right), the distance curve obtained with NMI presents several local extrema and is not symmetric. In contrast, the curve obtained with CCOD is symmetric with a global minimum around zero translation.

To assess the convergence of CCOD during pairwise registration, we performed an experiment, where a rigidly transformed moving binary image is registered to a fixed image. The original images, taken from the first synthetic dataset, are shown in Fig. 4, together with some intermediate registration results and the values of CCOD as cost function as the number of iteration increases. When misaligned, the value of CCOD is high. The better the alignment with the fixed image is, the smaller the value of CCOD becomes. Figure 5 shows the CCOD surface as a function of two out of the three free parameters of the rigid transformation. It can be seen that CCOD has a global minimum at the alignment of the images.

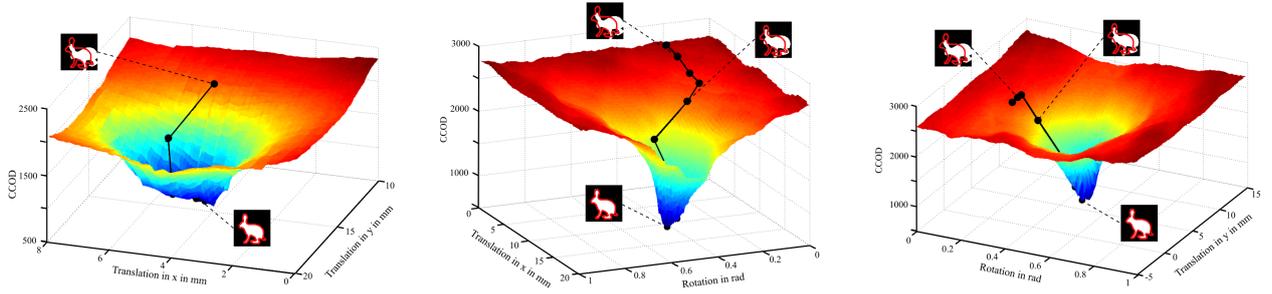


Figure 5: Registration example using CCOD on an image pair of the first synthetic dataset. The shape of CCOD as objective function is shown by varying in each plot two parameters of the rigid transformation (rotation in radians, translations in pixels). Left: Varying the translation in x and y; Middle: Varying the rotation angle and translation in x; Right: Varying the rotation angle and translation in y. The evolution with the iterations of the registration process is shown as black points.

6.5. Image registration

We evaluated the registration accuracy of CCOD and NMI/LNMI on the selected datasets using global (rigid) and local (B-Spline) transformation. For the rigid case, the transformations were randomly created with a maximal translation and rotation (specified individually for each dataset, see below). We deformed the moving image with these transformations and used their inverse as ground truth. For the local transformations, deformations were generated with a maximal pixel displacement using random displacements of the control points of a dense B-Spline grid. We created the corresponding deformation fields and, to ensure physically plausible transformations, smoothed them by a Gaussian kernel with standard deviation $\sigma = 8$. We used the smoothed deformation fields to deform the moving images and the smoothed inverse deformation fields as ground truth.

The registration accuracy was measured by the Root Mean Square Error (RMSE) between the estimated and ground truth transformation parameters. For rigid registration, we considered a 1 pixel or 1 mm error, respectively, equal to a 1° error to quantify the translational and rotational displacement from the ground truth in a single RMSE value. For both synthetic datasets, the accuracy is additionally measured by the Dice coefficient between the objects in the aligned and transformed moving image. The Dice coefficient between two sets A and B is given by $D = \frac{2|A \cap B|}{|A| + |B|}$ and measures the similarity of the two sets.

To test for significance of the results, we performed for all cases a Wilcoxon signed-rank test and reported the corresponding p-values. Significant differences in the accuracy of NMI/LNMI and CCOD at $p < 0.05$ are printed in bold.

6.5.1. Synthetic data

For the rigid case, 20 transformations were randomly generated with a maximal translation of ± 30 pixels and a maximal rotation of $\pm 86^\circ$. For the local transformations, a total of 10 deformations were generated with a maximal pixel displacement of ± 23 . In total, we performed 230 rigid and 130 non-rigid registration

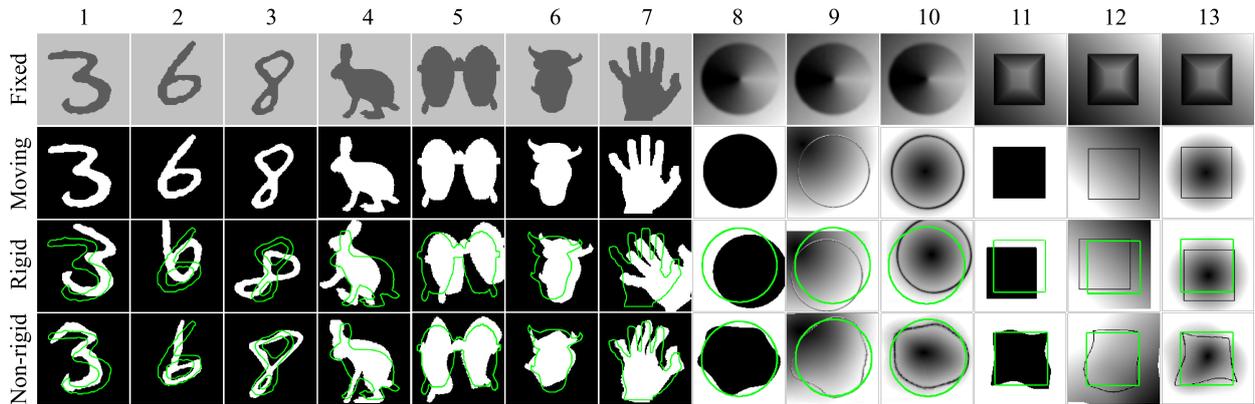


Figure 6: Synthetic datasets used for registration. From top to bottom: fixed images, moving images (aligned with fixed image) and (an example of) moving images deformed with rigid (third row from the top) and B-Spline (bottom row) transformation with the contour of the fixed image in green. First synthetic dataset: Column one to seven; second synthetic dataset: column eight to thirteen.

experiments for both NMI/LNMI and CCOD as distance measures. Figure 6 shows the image pairs of the synthetic datasets together with examples of initial rigid and non-rigid misalignments.

Table 1 shows the results. For the first synthetic dataset, both NMI/LNMI and CCOD were able to measure the distance between the images accurately. That was confirmed by the high Dice coefficients of the registration results. However, the registration using CCOD yielded higher Dice coefficients and significantly ($p < 0.001$) smaller RMSE with less standard deviation. For the second synthetic dataset, the registration results using CCOD were significantly more accurate ($p < 0.001$). On the first synthetic dataset with relatively large rigid deformations between the images, LNMI performed slightly worse than NMI. However, LNMI was designed to deal with locally varying intensity distributions in images, which are present in the second synthetic dataset. As expected, LNMI performed here much better than NMI, and slightly worse than CCOD. For the global non-rigid deformations, LNMI is comparable to CCOD.

6.5.2. Visual-infrared data

On the visual-infrared dataset, we performed pairwise rigid registration using the visual image as fixed image. The particular challenge of this dataset is that the background structures of the visual image are not seen in the infrared image, which shows only the pedestrian and the light source. In [65], it was reported that MI-based similarity measure yielded good results on this dataset. Here, we explored the accuracy of CCOD compared to NMI and, additionally, the robustness of both measures to the amount of initial displacement. We generated randomly two sets of rigid transformations (20 transformations for each set). For both sets, the maximal rotation was $\pm 45^\circ$ and the maximal translation ± 10 pixels for the first set and ± 30 pixels for the second set. In total, we performed 320 registration experiments both for NMI and CCOD as distance measure.

Table 1: Dice coefficients, mean registration errors (RMSE \pm standard deviation in pixels) and p-values for the synthetic datasets.

| | | First dataset | | Second dataset | |
|----------|------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| | | RMSE | Dice | RMSE | Dice |
| Rigid | NMI | 0.138 ± 0.090 | 0.987 ± 0.009 | 3.937 ± 3.150 | 0.904 ± 0.055 |
| | LNMI | 0.237 ± 0.123 | 0.986 ± 0.003 | 0.82 ± 0.886 | 0.982 ± 0.012 |
| | CCOD | 0.040 ± 0.043 | 0.993 ± 0.007 | 0.276 ± 0.160 | 0.994 ± 0.004 |
| | p | < 0.001 | | < 0.001 | |
| B-Spline | NMI | 1.254 ± 0.085 | 0.968 ± 0.014 | 1.776 ± 0.127 | 0.960 ± 0.010 |
| | LNMI | 1.168 ± 0.07 | 0.97 ± 0.004 | 1.63 ± 0.09 | 0.97 ± 0.012 |
| | CCOD | 1.169 ± 0.085 | 0.984 ± 0.008 | 1.631 ± 0.120 | 0.971 ± 0.006 |
| | p | < 0.001 | | < 0.001 | |

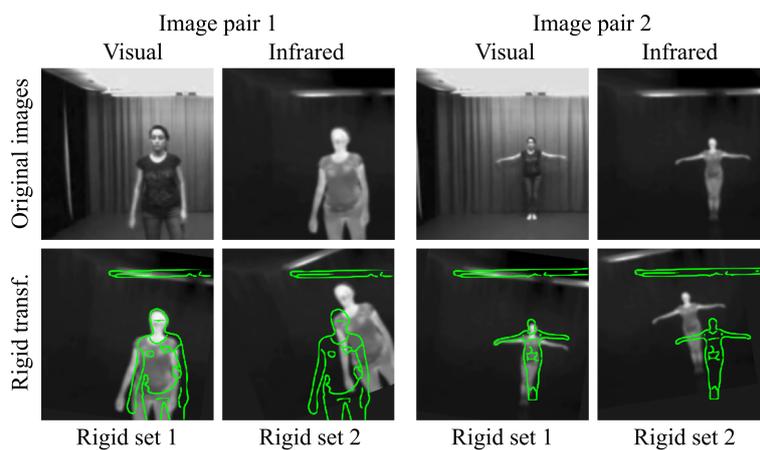


Figure 7: Top row: two image pairs of the visual-infrared dataset. Bottom row: examples of initial rigid displacements of the first rigid set and the second rigid set with the contours of the aligned image in green.

Table 2: Mean registration errors (RMSE \pm standard deviation in pixels) and p-values for the visual-infrared dataset on two sets of rigid transformation.

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Average |
|-------------|----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Rigid set 1 | NMI | 2.679 | 2.484 | 1.388 | 1.984 | 1.115 | 2.210 | 0.776 | 1.054 | 1.711 |
| | | ± 2.656 | ± 1.541 | ± 1.141 | ± 0.669 | ± 0.269 | ± 1.288 | ± 0.195 | ± 0.859 | ± 1.471 |
| | CCOD | 1.022 | 2.119 | 2.270 | 1.053 | 0.733 | 1.873 | 1.107 | 2.173 | 1.544 |
| | | ± 0.221 | ± 0.987 | ± 0.519 | ± 0.561 | ± 0.176 | ± 0.160 | ± 0.192 | ± 0.746 | ± 0.789 |
| | <i>p</i> | 0.011 | 0.430 | 0.003 | < 0.001 | < 0.001 | 0.261 | < 0.001 | 0.002 | 0.478 |
| Rigid set 2 | NMI | 5.41 | 7.169 | 2.592 | 2.784 | 2.365 | 4.160 | 2.172 | 3.882 | 3.816 |
| | | ± 3.45 | ± 2.437 | ± 2.260 | ± 0.958 | ± 2.687 | ± 2.329 | ± 2.107 | ± 3.452 | ± 3.044 |
| | CCOD | 1.34 | 2.734 | 3.822 | 1.623 | 1.285 | 2.295 | 1.468 | 3.224 | 2.224 |
| | | ± 0.39 | ± 1.260 | ± 1.918 | ± 0.772 | ± 0.917 | ± 0.814 | ± 0.502 | ± 1.158 | ± 1.393 |
| | <i>p</i> | < 0.001 | < 0.001 | 0.004 | < 0.001 | 0.090 | 0.007 | 0.784 | 0.546 | < 0.001 |

415 Figure 7 shows some examples of initial transformation of the two sets for two image pairs. Table 2 shows the results of pairwise registration using CCOD and NMI on both rigid transformation sets. Despite the missing background structures in the infrared images, CCOD yields low registration errors. As already reported in [65], MI-based distance measures perform well on this dataset and the accuracy of NMI and CCOD are comparable for the first rigid set of transformation. However, with increasing initial translation, 420 the error using NMI increases from 1.7 to 3.8 mm, while the error using CCOD only increases to 2.2 mm, yielding a significantly higher accuracy for CCOD over all image pairs ($p < 0.001$).

6.5.3. Clinical data

First clinical dataset. Intensity non-uniformity (INU) is a common artifact in MR imaging due to the inhomogeneity in the magnetic field and has an adverse effect on the performance of registration algorithms. 425 Despite of available methods to correct for such INU fields (see, e.g., [68]), registration techniques robust to those fields are advantageous. To test for this robustness, we added INU-fields (see Fig. 8) of various strength to the images prior registration. The field was different for the fixed and moving image.

We generated randomly a total of 30 rigid transformation with a ± 20 mm range for translation and $\pm 20^\circ$ range for rotation. Next, we created 10 non-rigid transformations with a maximum displacement of 430 ± 23 mm. We ensured (by controlling the positivity of the Jacobians) that the deformations were anatomically plausible. We performed in total 400 pairwise registrations (300 using rigid and 100 using B-Spline transformations), taking the T1 image as fixed image and the T2 as moving image in all cases. The results

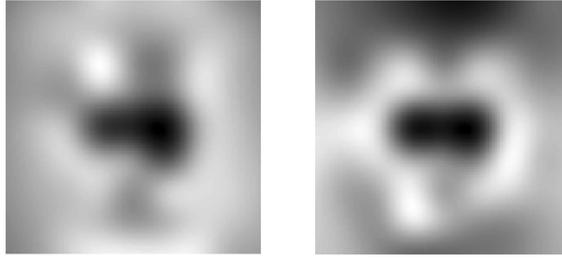


Figure 8: Example of intensity non-uniformity fields added to the T1-T2 dataset. The fields were extracted from real T2 weighted MR acquisitions using a 1.5T system (Aera, Siemens Healthcare).

Table 3: Mean rigid and B-Spline registration errors (RMSE \pm standard deviation in mm) and p-values for the T1-T2 dataset with various strengths of intensity non-uniformity (INU) fields.

| | | – | 16% | 24% | 31% | 40% |
|----------|------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Rigid | NMI | 0.124 ± 0.006 | 0.129 ± 0.008 | 0.132 ± 0.006 | 0.148 ± 0.057 | 0.157 ± 0.007 |
| | CCOD | 0.128 ± 0.005 | 0.129 ± 0.007 | 0.141 ± 0.008 | 0.141 ± 0.007 | 0.141 ± 0.007 |
| | p | 0.533 | 0.933 | 0.292 | 0.276 | 0.006 |
| B-Spline | NMI | 1.070 ± 0.016 | 1.078 ± 0.015 | 1.087 ± 0.014 | 1.106 ± 0.014 | 1.132 ± 0.014 |
| | CCOD | 1.051 ± 0.016 | 1.062 ± 0.015 | 1.073 ± 0.015 | 1.083 ± 0.015 | 1.095 ± 0.016 |
| | p | < 0.001 | < 0.001 | 0.004 | < 0.001 | < 0.001 |

for rigid and B-Spline registration are shown in Table 3 and Fig. 9. For rigid registration, the differences in accuracy between NMI and CCOD were non-significantly when adding no INU field and INU fields of strengths 16%, 24% and 31%. The errors for both measures grew with increasing strength of the INU field. However, the errors using CCOD increased slower and were, for strong INU fields, lower than the errors using NMI. In total, the error of NMI grew by 26% and the error of CCOD by 10%. For B-Spline registration, CCOD yielded significantly lower average RMSE for all cases (all $p \leq 0.002$). With increasing INU field, its error increased in total by 4.2% while the error of NMI increased by 5.8%. However, although CCOD yielded in some cases significantly better results, the differences in registration error between NMI and CCOD were in general very small and both measures performed comparably. Figure 9 shows registration examples of two patients using CCOD as distance measure. The initial rigid and non-rigid displacements were successfully recovered.

Second clinical dataset. The registration of US to MR images is a challenging task due to structural differences in the MR and US images. In the selected MR-US Brain dataset, the structures common to MR and US are the ventricular system and the tumor. Here, we explored the ability of CCOD to recognize those

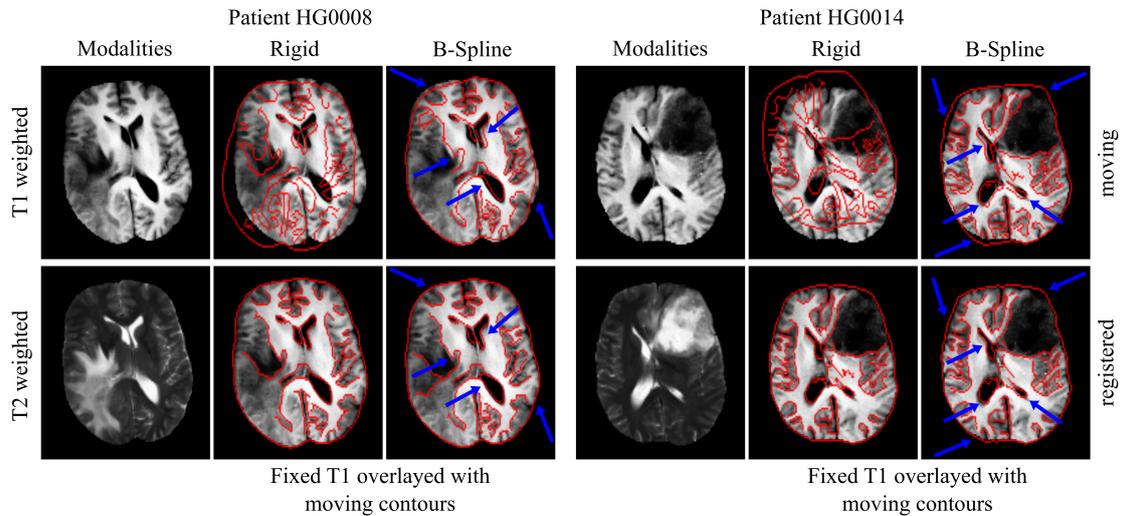


Figure 9: Examples of T2 to T1 registration for patient HG0008 and patient HG0014. Random rigid (top row, column two and five, respectively) and random B-Spline transformation (top row, column three and six, respectively); estimated transform for rigid (bottom row, column two and five) and B-Spline registration (bottom row, column three and six) using CCOD as distance measure. Blue arrows highlight regions where an initial misalignment was corrected by the registration.

structures in both images for the displacement estimation and we compared it to NMI.

We performed experiments in 3D and 2D. We created a 2D dataset by aligning the corresponding landmark sets (our ground truth). We extracted corresponding 2D slices in axial, coronal and sagittal direction of the MR and US volumes and transformed the US slices with randomly created transformations. We generated a total of 10 rigid transformation with a ± 10 mm range for translation and $\pm 10^\circ$ range for rotation to transform the US slices for the 2D dataset. We performed rigid registration to recover the random rigid displacements and local registration using B-Spline transformations to correct for the non-rigid deformations between pre-operative MR and intra-operative US. In total, we performed 200 registration experiments on the whole 2D dataset. Table 4 shows the mean RMSE of the transformation parameters for 2D rigid registration. We obtained a significantly lower RMSE (all $p < 0.001$) with less standard deviation using the proposed distance measure in comparison to NMI. The errors are around 1 mm, whereas the errors for NMI are around 2.6 mm in average. Figure 10 shows some examples of registration results for two patients using CCOD as distance measure. The rigid registration corrects for the randomly introduced translation and rotation of the US image towards the MR (Fig. 10, top and middle row). Further local registration using B-Splines improves the alignment (visually confirmed in Fig. 10, bottom row).

On the original 3D dataset, we performed rigid registration using CCOD and NMI as distance measures and we measured the registration accuracy by the average Euclidean distance between the corresponding landmarks (Fiducial Registration Error, FRE, in mm). Table 5 shows the FRE before registration (using the initial transformation provided with the database), and the FRE after rigid registration using CCOD

Table 4: Mean 2D rigid registration error (RMSE \pm standard deviation in mm) and p-values in the MR-US dataset.

| | 01 | 02 | 03 | 05 | 06 | 10 | 11 | 12 | 13 | 14 | Average |
|------|--------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| NMI | 2.646 | 2.578 | 2.286 | 2.701 | 2.642 | 2.355 | 3.122 | 2.138 | 2.246 | 2.984 | 2.590 |
| | ± 1.122 | ± 1.330 | ± 1.086 | ± 1.003 | ± 1.297 | ± 0.995 | ± 1.337 | ± 1.428 | ± 1.308 | ± 1.377 | ± 1.257 |
| CCOD | 2.108 | 1.294 | 0.898 | 1.100 | 0.988 | 0.828 | 0.798 | 0.607 | 0.907 | 0.976 | 1.050 |
| | ± 1.038 | ± 0.809 | ± 0.528 | ± 0.651 | ± 0.516 | ± 0.581 | ± 0.355 | ± 0.342 | ± 0.519 | ± 0.396 | ± 0.716 |
| p | 0.010 | < 0.001 |

Table 5: Euclidean distance (in mm) between Landmarks for 3D rigid registration in the MR-US dataset compared to results reported in the literature.

| | 02 | 03 | 05 | 06 | 10 | 11 | 12 | 13 | 14 | Average |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-----------------------------------|
| Initial | 6.30 | 9.38 | 2.62 | 2.30 | 2.99 | 1.52 | 3.70 | 5.15 | 3.77 | 4.27 ± 2.31 |
| NMI | 5.10 | 8.08 | 2.30 | 2.58 | 2.70 | 1.71 | 3.77 | 4.24 | 4.02 | 3.83 ± 1.22 |
| CCOD | 4.81 | 4.58 | 2.52 | 1.76 | 2.27 | 1.53 | 2.63 | 4.20 | 2.82 | 3.01 ± 1.22 |
| MIND [23] | 4.59 | 4.55 | 2.34 | 4.17 | 3.56 | 2.45 | 2.73 | 4.41 | 4.53 | 3.7 ± 0.96 |
| LMI [61] | 4.05 | 3.74 | 2.72 | 2.23 | 2.84 | 3.94 | 2.29 | 2.67 | 2.9 | 3.04 ± 0.69 |
| CoCoMI [69] | 3.22 | 3.03 | 2.18 | 2.2 | 2.04 | 2.48 | 2.16 | 2.64 | 2.07 | 2.45 ± 0.43 |
| SeSaMI [71] | 1.8 | 2.54 | 2.59 | 1.73 | 2.74 | 1.35 | 2.78 | 2.91 | 2.16 | 2.29 ± 0.55 |
| LC2 [70] | 1.64 | 2.43 | 2.26 | 2.2 | 2.09 | 1.76 | 2.45 | 3.71 | 2.76 | 2.37 ± 0.61 |

and NMI as distance measures. CCOD yields in average lower registration errors than NMI and improves the initial alignment.

The BITE database is used in the literature to evaluate registration methods developed for MR-US image alignment. In addition to our results on rigid registration, we report in Table 5 FRE errors published in [69, 70]. CCOD ranks average, comparing favorably to NMI, local mutual information (LMI, [61]) and modality independent neighborhood descriptor (MIND, [23]), but contextual conditioned mutual information (CoCoMI, [69]), linear correlation of linear combination (LC2, [70]) and self-similarity weighted mutual information (SeSaMI, [71]) have lower errors. This is not surprising since especially CoCoMI and LC2 are especially designed for MR-US registration. On the other hand, CCOD is a general measure, which outperforms other general measures (NMI, LMI, MIND) for the task of MR-US registration. SeSaMI, which uses not only intensity but also gradient information, performs best.

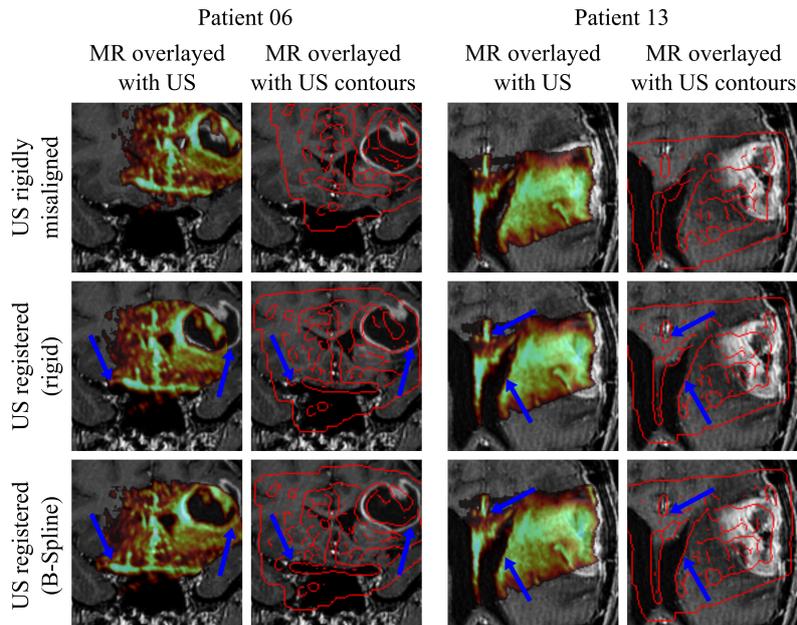


Figure 10: Examples of US to MR registration of patient 06 (left) and 13 (right). Random rigid transformation (top row); estimated transform for rigid (middle row) and B-Spline registration (bottom row) using CCOD as distance measure. Blue arrows highlight regions where an initial misalignment was corrected by the registration.

6.6. Computational complexity

In this work, our baseline for comparison is NMI, a de facto state-of-the-art similarity measure for image registration. The complexity of NMI is dependent on the MI estimation approach. The implementation we used is based on [72] and the complexity is $\mathcal{O}(sm^2p)$ [73] (with derivative estimation), where s is the number of samples, m the number of histogram bins and p the number of transform parameters.

The complexity of CCOD is $\mathcal{O}(n^3qp)$ [9], meaning that it scales linearly with the patch size q and the number of transformation parameters p . It scales cubic with n , the number of rows of the image graph Laplacian. As an illustration, Fig. 11 shows the computational time when varying those three parameters in a toy example (using an image pair of the first synthetic dataset).

7. Discussion and conclusions

The fusion and combination of images from multiple modalities is important in many applications. Typically, this process consists of the alignment of the images and the combination of the complementary information. In this work, we focused on the former part and, to this end, we proposed a new multimodal distance measure for image registration based on the commutativity of image graph Laplacians.

We build upon the recent framework of simultaneous diagonalization [9] and its equivalent formulation of the closest commuting Laplacians [5] to define a metric between multimodal images accounting for how

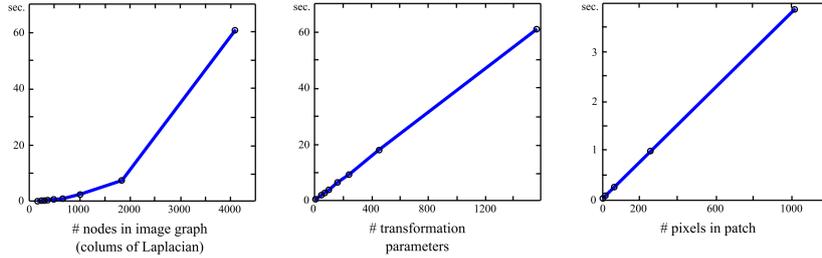


Figure 11: Analysis of the computational complexity of CCOD. Average time of CCOD evaluation when varying (a): the number of nodes in the image graph, with three free transformation parameters and image size 128×128 ; (b) the number of transformation parameters, with 256 nodes and image size 128×128 ; (c) the image size (number of pixels per patch), with three transformation parameters and 256 nodes.

well the image Laplacians are simultaneously diagonalizable. We express the metric and its derivatives with respect to the image transformation and thus perform optimization over these parameters. As in [5], we employ the fact that almost commuting matrices are close to commuting matrices, which in turn, are simultaneously diagonalizable. We use this mathematical relationship to use Laplacian commutativity as a criterion for image structure similarity. In contrast to [5], we do not use the eigendecomposition of the closest commuting Laplacians for our analysis. We use the image Laplacian matrix to capture the structural information and we consider that, if two images are aligned, then their eigenvectors are similar or, equivalently, they will almost commute. Therefore, the main contribution is the adaptation and application of the mathematical framework presented in [5] to the problem of image registration and to show how this approach successfully aligns images of multiple and challenging datasets.

Spectral methods have become popular techniques in several applications but have received limited attention in dense image registration. In [6, 8, 10], the similarity of the spectral coordinates of two image graph Laplacians was used to include the geometric structure into the registration problem. However, the method in [6] is only applicable to monomodal images, because of the incompatibility of eigenspaces of non-isometric manifolds, as explained in Section 1. In [8, 10] the first embedding coordinate of Laplacian eigenmaps and diffusion maps, respectively, were used, with suitable methods to make the eigenspaces comparable, for multimodal image registration. Since it is computationally very expensive to compute the spectral embeddings in each iteration, the embedding coordinates were used as a new representation of the images and those representations were registered. However, in general one cannot assume that the transformation of a spectral representation of an image is the same as the spectral representation of a transformed image.

We addressed several of the aforementioned limitations. First, CCOD is applicable to multimodal images while transforming directly the image instead of their spectral representations without realignment of the eigenspaces. Second, the explicit computation of the spectral coordinates is avoided by using the relation

between simultaneous diagonalization and commutativity of matrices as a criterion for image structure preservation.

The formulation of CCOD is in spirit of the Laplacian colormaps presented in [13, 14]. However, our
520 framework and application show substantial differences. Eynard et al. transform the color space of an image
by interpreting the structural similarity in terms of color and contrast. Our work, on the other hand, focuses
on the geometry of two multimodal images. Using CCOD as distance measure, the registration is guided
such that the geometrical structure of a transformed image is most similar to a given image. By doing so, we
have to consider not only linear transformations (as linear color mapping and spatial rigid transformations)
525 but also higher-order spatial transformation, such as B-Splines, which requires the optimization of a higher
number of parameters (in the order of $\gg 100$). In addition, we introduced patch-based graph Laplacians to
be able to deal with high resolution 2D and 3D images. Additionally, to account for the different range of
graph Laplacians of different imaging modalities, we considered a normalized version of the graph Laplacian,
the random-walk Laplacian, and provided the corresponding derivatives of CCOD in Appendix B.

We tested the performance of the proposed image distance measure on very challenging both synthetic
530 and real multimodal datasets. Results confirm that enforcing the commutativity of the graph Laplacians
favorably compares to NMI, a state-of-the-art measure for multimodal image registration. We tested the
proposed distance measure on datasets where NMI was expected to have a good performance (first synthetic
dataset, visual-infrared dataset and T1-T2 dataset) and on datasets where NMI was expected to have a
535 poor performance (second synthetic dataset and MR-US dataset) due to the presence of intensity gradients
or very different intensity distributions. NMI assumes a global statistical relationship between intensities
and is therefore not suited for the registration of these datasets. On the first type of dataset (first syn-
thetic, visual-infrared and T1-T2) CCOD performed comparable to NMI. The improvements of CCOD with
respect to NMI are not very high. However, CCOD showed higher robustness to larger initial displace-
540 ments (visual-infrared). Despite the missing background structures in the infrared images, CCOD yields
low registration errors on the visual-infrared images. On the other datasets (second synthetic and MR-US),
CCOD demonstrated a substantial increase in accuracy over NMI which, as expected, performed poorly
even for simple rigid transformations (see, e.g., the high RMSE and high variance for rigid registration in
Table 1 and Tables 4, and 5 for MR-US registration). On these two datasets, where NMI would not be a
545 good choice as distance measure, we additionally included results obtained using LNMI (second synthetic
dataset) and results reported in the literature (MR-US dataset). LNMI (LMI for MR-US data), designed
to address some of the short-comings of NMI, outperformed NMI on these datasets, as expected. CCOD
obtained in most of the cases lower registration errors than LNMI. Compared to the results reported in the
literature, CCOD performed average on the MR-US dataset: it yielded better results than other general
550 measures (NMI, LMI, MIND) but it did not outperform specific MR-US distance measures (CoCoMI, LC2).
Self-similarity weighted mutual information (SeSaMI, [71]), which uses not only intensity but also gradient

information, performed best.

It is known that topological changes in a graph (e.g. adding or removing edges) can significantly perturb the eigenspace of the graph Laplacian. Thus, conventional spectral matching cannot handle missing data (i.e., different structures in the images to register), since they produce two independent sets of eigenvectors that cannot be matched. In [5], the closest commuting operator is shown to be robust to topological noise and produce single consistent representation. In this regard, from our experiments, we conclude that our method can handle well: (i) missing correspondences between structures (as shown in Section 6.5.2 with the infrared-visual data); (ii) different intensity distributions (as shown in Section 6.5.1 with the synthetic dataset 2 and in Section 6.5.3 with the second clinical dataset); and (iii) non-uniform intensity distributions (as shown in Section 6.5.3 with the first clinical dataset). As long as a main structure is present in both modalities, CCOD was robust to those missing objects and different (and possibly non-uniform) intensity distributions.

CCOD is a rather global measure. Our results indicate that this is especially beneficial for large global deformations (Tables 1, 2, 4, 5). In the case of smaller initial deformations (Tables 3), registration errors were lower and differences between NMI and CCOD are less relevant. Note that our approach for generating the non-rigid deformations in Section 6.5.3 is only applicable to relatively small deformations. For larger deformations, another procedure (for example as in [6]) should be employed to ensure anatomic plausibility.

We used NMI as state-of-the-art multimodal distance measure for comparison. Despite its several extensions to deal with its limitations (e.g., [74]), NMI has become the de facto standard for multimodal registration, being one of the most successful multimodal distance measures and widely used in practice (see, e.g., the studies in [75, 76]). When NMI fails, e.g., for MR-US registration, the distance measures that have been proposed are rather specific for the application at hand. In the case of the very challenging MR-US registration problem, one approach is to simulate an US image from the segmented MR image and align the original US with the simulation [77]. Other extensions of NMI, as local NMI [61, 74] and phase NMI [64], or measures using image gradients [19], have shown to address successfully some shortcomings of the original NMI. However, our objective was not to compare extensively different distance measures for image registration or to find the best measure for all datasets, but to show that by using the commutativity of image graph Laplacians, multimodal image structures can be successfully compared. With the proposed CCOD, we presented a context-free, non-specific measure with good performance on very different and challenging multimodal datasets.

Amongst other data, we performed experiments on clinical data (MR and US brain images). Aligning MR and US images in a well known and challenging problem, both for rigid and non-rigid deformations. Regarding the T1-T2 MR data, the experiments using rigid transformations between images of the same patient are realistic scenarios, because T1 and T2 scans are in general not aligned. In contrast, the experiments using B-spline transformations represent a less common scenario. Non-rigid distortions of brain

images of the same patient may occur in follow-up studies (e.g., after tumor resection or for monitoring neurodegenerative diseases such as Alzheimer’s). However, the evaluation of registration methods becomes more difficult when using real data of either follow-up studies of the same patients or cross-sectional studies analyzing changes of different patients, because of the lack of ground truth. Therefore, we used a synthetic registration scenario, as explained in Section 6.5.3. This allows us to assess the performance of CCOD on clinical data with rich texture and local non-rigid distortions, but with a reliable evaluation using the ground truth transformations.

The runtime of CCOD is, especially for 3D registration, its main limitation. The main influence have the number of nodes used to construct the graph Laplacians (cubic scale) and the number of transform parameters (in particular for B-spline registration, where the number of control points can be very high). To reduce computation time, we used patch-based image graphs instead of pixel/voxel-based graphs to construct the Laplacians and to reduce their size. The size of the patches increases in higher resolution levels. This may imply some information lost for detailed structures, resulting in less accurate registrations, in particular for local transformations. Comparing the results of the two synthetic datasets, we can see that the RMSE values are higher for the second synthetic dataset. The main difference between the two datasets is that the first one has only two different intensity values whereas intensity gradients, interpretable as texture patterns, are present in the second one.

To improve the registration of images with texture, the proposed distance could be combined with a suitable intensity-based measure that captures detailed structures. A similar strategy was followed in [6], where the Log-Demons framework was used in combination with the comparison of the individual eigenvectors of the images graph Laplacians. Moreover, the Laplacian weights in Eq. (5) could be computed using other type of information that could potentially improve the matching accuracy.

For the experiments on the clinical datasets, we did not use a combination of distance measures, as we suggested above. In this work, it was not the objective to find the best distance measure for T1-T2 or MR-US registration. Instead, we showed that the comparison of image graph Laplacians is a suitable criterion for multimodal registration of medical images. For future work, we plan to improve the registration accuracy by using a combination of CCOD with a state-of-the-art distance measure, designed, e.g., especially for MR-US registration.

A limitation of CCOD is that many parameters have to be selected (σ , β , k and the size of the patches for the patch-based image graph). The scale parameters can be explained in light of the scale-space theory in [60]. The parameter σ_1 can be interpreted as the tonal scale (intensity scale or resolution in the domain of the sampled values), while the parameter σ_2 can be interpreted as the inner scale (scale or resolution at which the image is observed). The patch size N would be a scope parameter (outer scale) that describes the spatial extent of the region of interest. The interplay between these parameters is given by n -the number of patches- and ν -related to the effective extent of the patch. Both σ_1 and σ_2 are adaptively computed

from the data using n and ν to characterize the structure up to an overall scaling and size. Our parameter β corrects for scale differences between the intensities and spatial coordinates, but could be “absorbed” by σ_2 by setting $\sigma_2 \leftarrow \sigma_2/\sqrt{\beta}$. The performance of the measure is particularly sensitive to the choice of σ_i , $i = 1, 2$, in Eq. (5). The sensitivity of the parameter σ for the representativity of the image graph was already mentioned in [37]. To overcome this problem, we provided automatic estimations for σ and β using Eqs. (8) and (9), respectively.

Even though the eigenspaces are not computed explicitly, the proposed similarity measure is relatively computationally expensive. Note that the most computational expensive part is not the storage of the Laplacians itself, but the computation of the derivatives of the cost function when many parameters have to be optimized. This is not problematic for rigid transformations, where there are only three parameters in 2D and six in 3D. However, for B-spline transformations, where hundreds of parameters need to be optimized, the computation time increases with the number of control points. We addressed this problem by working with patch-based image graphs instead of voxel-based. Another solution would be to parallelize the code and possibly to subdivide the graph Laplacian for faster computation times.

We have formulated the registration in an asymmetric way (the cost function is asymmetric), but CCOD could eventually be used in a symmetric registration framework (e.g. by transforming both images simultaneously or by evaluating the distance on both images domains). Also, the framework of the CCO’s [5] is extendable to more than two images, which is interesting for groupwise registration and template construction. Simultaneous diagonalization has already been used for more than two Laplacians. In [78], e.g., the coupled joint eigenspace of brain networks was used for longitudinal connectivity analysis. For group-wise image registration, the overall aim would be to transform all images at the same time, such that their graph Laplacians all have approximately the same eigenvectors. In our setting, one approach could be to transform simultaneously all images such that pairwise image Laplacians almost commute. This could be measured as the average of the CCO’s of all image pairs.

The process of image fusion consists of the alignment of the images and the combination of the complementary information. In this work, we used the framework of the CCO’s for image alignment. However, the proposed image distance CCOD could also be used to create the fused image (by minimizing the distance between sources and fused image, similar to [13]) or as a measure of quality assessment in image fusion (i.e., the lower the CCOD between sources and fused image, the higher the quality).

To conclude, we have presented a new approach to measure image similarity for image registration and we have shown that the proposed distance measure based on Laplacian commutativity successfully aligns images. It can handle different modalities and is robust to intensity non-uniformity and large initial deformations, which is important for the combination of very different, complementary images. Thus, we believe that it has a high potential for multimodal image registration and fusion.

Appendix A. Asymptotic Metric Property

Here we look at the metric properties of the proposed distance measure CCOD in Eq. (7).

A metric on the set of real matrices $\mathbb{R}^{n \times n}$ is a function $d : \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$, satisfying the following properties for any $A, B, C \in \mathbb{R}^{n \times n}$:

- (P1) $d(A, B) \geq 0$ *(non-negativity)*
- (P2) $d(A, B) = 0$ iff $A = B$ *(identity)*
- (P3) $d(A, B) = d(B, A)$ *(symmetry)*
- (P4) $d(A, C) \leq d(A, B) + d(B, C)$ *(triangle-inequality)*

If (P2) holds only in one direction (i.e., $d(A, B) = 0$ for some $A \neq B$), then function d is a pseudo-metric. If d satisfies (P1)-(P3) but not (P4), d is called a semi-metric.

One can easily show that the Frobenius distance $\|A - B\|_F$ is a metric, while the Frobenius norm of the commutator $\|AB - BA\|_F$ is not a metric, since neither (P2) nor (P4) holds. However, it can be shown [56] that the norm of the commutator is asymptotically a pseudo-metric in the sense that it satisfies (P4) asymptotically for large matrices drawn from a “good” distribution (e.g., i.i.d. with uniform, normal or Rademacher distribution). Therefore, we conclude that the proposed distance measure in Eq. (7) is asymptotically a pseudo-metric.

In the image registration problem, we presume that, if the images are not aligned, their structure is different. Moreover, in practice, the fixed and the moving image have not, even after registration, exactly the same image structure. Hence, their graph Laplacians do not fully commute and (P2) holds. This allows us to conclude that the distance measure defined in Eq. (7) is asymptotically a metric.

Appendix B. Derivatives

For the optimization of the proposed distance measure CCOD in Eq. (7), the derivatives with respect to the transformation parameters \mathbf{p} have to be computed:

$$\frac{d}{d\mathbf{p}} \mathcal{D} = \alpha \frac{d}{d\mathbf{p}} \|\mathbf{L}^F \mathbf{L}^M(\mathbf{p}) - \mathbf{L}^M(\mathbf{p}) \mathbf{L}^F\|_F^2 + (1 - \alpha) \frac{d}{d\mathbf{p}} \|\mathbf{L}^F - \mathbf{L}^M\|_F^2. \quad (\text{B.1})$$

More details on the computation of the derivatives can be found in the supplementary material.

For simplicity of notation, we suppress the superscript M for the moving image I^M , the moving graph Laplacian \mathbf{L}^M and corresponding weights w_{ij}^M . The edge weights of the moving image graph Laplacian can then be written as:

$$w_{ij}(\mathbf{p}) = \exp\left(-\left[\frac{g(\mathbf{p})}{2\sigma_1^2} + \beta \frac{\|u_i - u_j\|_2^2}{2\sigma_2^2}\right]\right) \quad (\text{B.2})$$

with $g : \mathbb{R}^p \rightarrow \mathbb{R}$,

$$g(\mathbf{p}) = \|\mathbf{I}(\phi(\mathbf{p}, u_i)) - \mathbf{I}(\phi(\mathbf{p}, u_j))\|^2 \quad (\text{B.3})$$

675 and $\mathbf{I}(\phi(\mathbf{p}, u_i)) = (I \circ \phi(\mathbf{p}, u_i^1), \dots, I \circ \phi(\mathbf{p}, u_i^N))$.

We can then rewrite Eq. (B.1) as

$$\frac{d\mathcal{D}}{d\mathbf{p}} = \left(\alpha \frac{d}{d\mathbf{L}} \mathcal{D}_1 + (1 - \alpha) \frac{d}{d\mathbf{L}} \mathcal{D}_2 \right) \frac{d}{dw} \mathbf{L} \frac{d}{d\mathbf{p}} w$$

where

$$\begin{aligned} \mathcal{D}_1(\mathbf{L}(w(\mathbf{p}))) &= \| \mathbf{L}^F \mathbf{L}(w(\mathbf{p})) - \mathbf{L}(w(\mathbf{p})) \mathbf{L}^F \|_F^2, \\ \mathcal{D}_2(\mathbf{L}(w(\mathbf{p}))) &= \| \mathbf{L}^F - \mathbf{L}(w(\mathbf{p})) \|_F^2. \end{aligned}$$

Derivatives of the similarity term w.r.t. the edge weights

We determine the derivatives of

$$\frac{d}{d\mathbf{L}} \mathcal{D}_k \frac{d}{dw} \mathbf{L}(w) = \left[\frac{d}{d\mathbf{L}} \mathcal{D}_k \frac{\partial}{\partial w_{ij}} \mathbf{L}(w) \right]_{\substack{i,j=1,\dots,n \\ i < j}}$$

for $k = 1, 2$ and for the unnormalized and normalized graph Laplacian. For simplicity, we write $l_{ij}(w) = l_{ij}$.

1) Unnormalized graph Laplacian:

$$\mathbf{L} = \mathbf{L}_u = \mathbf{D} - \mathbf{W}$$

(a) for $k = 1$:

$$\begin{aligned} \frac{\partial}{\partial w_{ij}} \mathcal{D}_1(\mathbf{L}) &= \frac{\partial}{\partial w_{ij}} \| \mathbf{L}(w) \mathbf{L}^F - \mathbf{L}^F \mathbf{L}(w) \|_F^2 \\ &= 2 \left((\mathbf{O}_2 - ((\mathbf{L}(w) \mathbf{L}^F - \mathbf{L}^F \mathbf{L}(w)) (\mathbf{L}^F)^T - \mathbf{O}_1 + (\mathbf{L}^F)^T (\mathbf{L}(w) \mathbf{L}^F - \mathbf{L}^F \mathbf{L}(w)))) \right)_{ij} \end{aligned}$$

with superindex ‘T’ denoting transpose,

$$\mathbf{O}_1 = (\text{diag}((\mathbf{L}^F)^T (\mathbf{L}(w) \mathbf{L}^F - \mathbf{L}^F \mathbf{L}(w))), \dots, \text{diag}((\mathbf{L}^F)^T (\mathbf{L}(w) \mathbf{L}^F - \mathbf{L}^F \mathbf{L}(w))))$$
 and

$$\mathbf{O}_2 = (\text{diag}((\mathbf{L}(w) \mathbf{L}^F - \mathbf{L}^F \mathbf{L}(w)) (\mathbf{L}^F)^T), \dots, \text{diag}((\mathbf{L}(w) \mathbf{L}^F - \mathbf{L}^F \mathbf{L}(w)) (\mathbf{L}^F)^T)).$$

(b) for $k = 2$:

$$\frac{\partial}{\partial w_{ij}} \mathcal{D}_2(\mathbf{L}(w)) = \frac{\partial}{\partial w_{ij}} \| \mathbf{L}^F - \mathbf{L}(w) \|_F^2 = 2 (\mathbf{O} + \mathbf{L}^F - \mathbf{L}(w))_{ij}$$

$$\text{with } \mathbf{O} = (\text{diag}(\mathbf{L}(w) - \mathbf{L}^F), \dots, \text{diag}(\mathbf{L}(w) - \mathbf{L}^F)) \in \mathbb{R}^{n,n}.$$

680 2) Normalized graph Laplacian:

$$\mathbf{L} = \mathbf{L}_{rw} = \mathbf{D}^{-\frac{1}{2}} \mathbf{L}_u = \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W})$$

(a) for $k = 1$:

$$\frac{\partial}{\partial w_{ij}} \mathcal{D}_1(\mathbf{L}) = \frac{\partial}{\partial w_{ij}} \| \mathbf{L}(w) \mathbf{L}^F - \mathbf{L}^F \mathbf{L}(w) \|_F^2$$

$$= 2 \left(\frac{1}{\mathbf{O}_3^2} \circ (-\mathbf{O}_3 \circ (\mathbf{L}(w)\mathbf{L}_F - \mathbf{L}_F\mathbf{L}(w))\mathbf{L}_F^T + \mathbf{W}(w) \circ (\mathbf{L}(w)\mathbf{L}_F - \mathbf{L}_F\mathbf{L}(w))\mathbf{L}_F^T + \mathbf{O}_4 + \mathbf{O}_3 \circ (\mathbf{L}_F^T(\mathbf{L}(w)\mathbf{L}_F - \mathbf{L}_F\mathbf{L}(w))) - \mathbf{W}(w) \circ (\mathbf{L}_F^T(\mathbf{L}(w)\mathbf{L}_F - \mathbf{L}_F\mathbf{L}(w))) - \mathbf{O}_5) \right)_{ij}$$

with \circ denoting the *Hadamard product* between two matrices,

$$\mathbf{O}_3 = (\text{diag}(\mathbf{D}), \dots, \text{diag}(\mathbf{D})) \in \mathbb{R}^{n,n},$$

$$\mathbf{O}_4 = (\text{diag}((\mathbf{L}(w)\mathbf{L}_F - \mathbf{L}_F\mathbf{L}(w))\mathbf{L}_F^T\mathbf{W}(w)^T), \dots, \text{diag}((\mathbf{L}(w)\mathbf{L}_F - \mathbf{L}_F\mathbf{L}(w))\mathbf{L}_F^T\mathbf{W}(w)^T)) \in \mathbb{R}^{n,n},$$

$$\mathbf{O}_5 = (\text{diag}(\mathbf{L}_F^T(\mathbf{L}(w)\mathbf{L}_F - \mathbf{L}_F\mathbf{L}(w))\mathbf{W}(w)^T), \dots, \text{diag}(\mathbf{L}_F^T(\mathbf{L}(w)\mathbf{L}_F - \mathbf{L}_F\mathbf{L}(w))\mathbf{W}(w)^T)) \in \mathbb{R}^{n,n}.$$

(b) for $k = 2$:

$$\begin{aligned} \frac{\partial}{\partial w_{ij}} \mathcal{D}_2(\mathbf{L}(w)) &= \frac{\partial}{\partial w_{ij}} \|\mathbf{L}^F - \mathbf{L}(w)\|_F^2 \\ &= \left(\frac{2}{\mathbf{O}_3^2} \circ (\mathbf{O}_3 \circ (\mathbf{L}^F - \mathbf{L}(w)) - \mathbf{W}(w) \circ (\mathbf{L}^F - \mathbf{L}(w)) - \mathbf{O}_6) \right)_{ij} \end{aligned}$$

$$\text{with } \mathbf{O}_3 = (\text{diag}(\mathbf{D}), \dots, \text{diag}(\mathbf{D})) \in \mathbb{R}^{n,n}$$

$$\text{and } \mathbf{O}_6 = (\text{diag}((\mathbf{L}^F - \mathbf{L}(w))\mathbf{W}(w)^T), \dots, \text{diag}((\mathbf{L}^F - \mathbf{L}(w))\mathbf{W}(w)^T)) \in \mathbb{R}^{n,n}.$$

Acknowledgements. V. A. Zimmer is supported by the grant FI-DGR 2013 (2013 FI B00159) from the Generalitat de Catalunya. This research was partially funded by the Spanish Ministry of Economy and Competitiveness (TIN2012-35874).

References

- [1] B. Zitová, J. Flusser, Image registration methods: a survey, *Image and Vision Computing* 21 (11) (2003) 977 – 1000.
- [2] T. Stathaki, *Image fusion: algorithms and applications*, Academic Press, 2011.
- [3] A. P. James, B. V. Dasarathy, Medical image fusion: A survey of the state of the art, *Information Fusion* 19 (2014) 4 – 19, special Issue on Information Fusion in Medical Image Computing and Systems.
- [4] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, *Information Fusion* 4 (4) (2003) 259 – 280.
- [5] M. M. Bronstein, K. Glashoff, T. A. Loring, Making Laplacians commute, *CoRR* arXiv:1307.6549.
- [6] H. Lombaert, L. Grady, X. Pennec, N. Ayache, F. Chériet, Spectral log-demons: Diffeomorphic image registration with very large deformations, *International Journal of Computer Vision* 107 (3) (2014) 254–271.
- [7] H. Lombaert, L. Grady, J. R. Polimeni, F. Chériet, Focusr: Feature oriented correspondence using spectral regularization—a method for precise surface matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (9) (2013) 2143–2160.
- [8] G. Piella, Diffusion maps for multimodal registration, *Sensors* 14 (6) (2014) 10562–10577.
- [9] D. Eynard, A. Kovnatsky, M. Bronstein, K. Glashoff, A. Bronstein, Multimodal manifold analysis by simultaneous diagonalization of Laplacians, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 37 (12) (2015) 2505–2517.

- [10] C. Wachinger, N. Navab, Entropy and Laplacian images: Structural representations for multi-modal registration, *Medical Image Analysis* 16 (1) (2012) 1–17.
- [11] R. R. Coifman, M. J. Hirn, Diffusion maps for changing data, *Applied and Computational Harmonic Analysis* 36 (1) (2013) 79–107.
- 705 [12] D. Eynard, K. Glashoff, M. M. Bronstein, A. M. Bronstein, Multimodal diffusion geometry by joint diagonalization of Laplacians, arXiv:1209.2295.
- [13] D. Eynard, A. Kovnatsky, M. M. Bronstein, Laplacian colormaps: a framework for structure-preserving color transformations, *Computer Graphics Forum* 33 (2) (2014) 215–224.
- [14] A. Kovnatsky, D. Eynard, M. M. Bronstein, Gamut mapping with image Laplacian commutators, *IEEE Int. Conf. on Image Processing* (2014) 635–639.
- 710 [15] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: A survey, *IEEE Transactions on Medical Imaging* 32 (7) (2013) 1153–1190.
- [16] M. A. González-Ballester, A. Zisserman, M. Brady, Segmentation and measurement of brain structures in MRI including confidence bounds, *Medical Image Analysis* 4 (3) (2000) 189 – 200.
- 715 [17] C. Studholme, D. Hill, D. Hawkes, An overlap invariant entropy measure of 3D medical image alignment, *Pattern Recognition* 32 (1) (1999) 71–86.
- [18] M. Bansal, K. Daniilidis, Joint spectral correspondence for disparate image matching, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2013, pp. 2802–2809.
- [19] E. Haber, J. Modersitzki, Intensity gradient based registration and fusion of multi-modal images, in: *Medical Image Computing and Computer-Assisted Intervention*, Vol. 4191 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2006, pp. 726–733.
- 720 [20] M. P. Heinrich, M. Jenkinson, F. V. Gleeson, S. M. Brady, J. A. Schnabel, Deformable multimodal registration with gradient orientation based on structure tensor, *Annals of the British Machine Vision Association* 5 (2011) 1–11.
- [21] D. D. Nigris, D. L. Collins, T. Arbel, Multi-modal image registration based on gradient orientations of minimal uncertainty, *IEEE Transactions on Medical Imaging* 31 (12) (2012) 2343–2354.
- 725 [22] E. Shechtman, M. Irani, Matching local self-similarities across images and videos, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [23] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, S. M. Brady, J. A. Schnabel, Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration, *Medical Image Analysis* 16 (7) (2012) 1423 – 1435.
- 730 [24] M. P. Heinrich, M. Jenkinson, B. W. Papie, S. M. Brady, J. A. Schnabel, Towards realtime multimodal fusion for image-guided interventions using self-similarities, in: *Medical Image Computing and Computer-Assisted Intervention*, Vol. 8149 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2013, pp. 187–194.
- [25] P. Aljabar, R. Wolz, D. Rueckert, Manifold learning for medical image registration, segmentation, and classification, in: K. Suzuki (Ed.), *Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis*, IGI Global, Hershey, PA, USA, 2012, pp. 351–372.
- 735 [26] K. Pearson, On lines and planes of closest fit to systems of points in space, *Philosophical Magazine* 2 (6) (1901) 559–572.
- [27] J. B. Tenenbaum, V. d. Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science* 290 (5500) (2000) 2319–2323.
- 740 [28] A. Xu, X. Jin, P. Guo, Two-dimensional PCA combined with PCA for neural network based image registration, in: *Advances in Natural Computation*, Vol. 4222 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2006, pp. 696–705.
- [29] A. Xu, P. Guo, Isomap and neural networks based image registration scheme, in: *Advances in Neural Networks*, Vol. 3972

of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2006, pp. 486–491.

- 745 [30] M. Hein, J.-Y. Audibert, U. v. Luxburg, Graph laplacians and their convergence on random neighborhood graphs, *Journal of Machine Learning Research* 8 (Jun) (2007) 1325–1368.
- [31] M. Reuter, S. Biasotti, D. Giorgi, G. Patanè, M. Spagnuolo, Discrete Laplace–Beltrami operators for shape analysis and segmentation, *Computers & Graphics* 33 (3) (2009) 381–390.
- [32] U. von Luxburg, A tutorial on spectral clustering, *Statistics and Computing* 17 (4) (2007) 395–416.
- 750 [33] B. Vallet, B. Lévy, Spectral geometry processing with manifold harmonics, *Computer Graphics Forum* 27 (2) (2008) 251–260.
- [34] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, L. Guibas, Functional maps: A flexible representation of maps between shapes, *ACM Transactions on Graphics* 31 (4) (2012) 30:1–30:11.
- [35] Y. Shi, R. Lai, D. J. J. Wang, D. Pelletier, D. Mohr, N. Sicotte, A. W. Toga, Metric optimization for surface analysis in the Laplace–Beltrami embedding space, *IEEE Transactions on Medical Imaging* 33 (7) (2014) 1447–1463.
- 755 [36] H. Lombaert, M. Arcaio, N. Ayache, Brain transfer: Spectral analysis of cortical surfaces and functional maps, in: *Information Processing in Medical Imaging*, Vol. 9123 of Lecture Notes in Computer Science, Springer International Publishing, 2015, pp. 474–487.
- [37] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Computation* 15 (6) (2003) 1373–1396.
- 760 [38] R. R. Coifman, S. Lafon, Diffusion maps, *Applied and Computational Harmonic Analysis* 21 (1) (2006) 5 – 30.
- [39] T. Vercauteren, X. Pennec, A. Perchant, N. Ayache, Symmetric log-domain diffeomorphic registration: A demons-based approach, in: *Medical Image Computing and Computer-Assisted Intervention*, Vol. 5241 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2008, pp. 754–761.
- 765 [40] H. Lombaert, J. Sporring, K. Siddiqi, Diffeomorphic spectral matching of cortical surfaces, in: *Information Processing in Medical Imaging*, Vol. 7917 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 376–389.
- [41] M. Reuter, F.-E. Wolter, N. Peinecke, Laplace–Beltrami spectra as ‘Shape-DNA’ of surfaces and solids, *Computer-Aided Design* 38 (4) (2006) 342–366.
- [42] B. Lévy, Laplace–Beltrami eigenfunctions towards an algorithm that “understands” geometry, in: *IEEE Int Conference on Shape Modeling and Applications*, 2006, pp. 13–13.
- 770 [43] V. A. Zimmer, B. Glocker, N. Hahner, E. Eixarch, G. Sanromà, E. Gratacós, D. Rueckert, M. A. G. Ballester, G. Piella, Learning and combining image neighborhoods using random forests for neonatal brain disease classification, *Medical Image Analysis* 42 (2017) 189–199.
- [44] V. A. Zimmer, K. Lekadir, C. Hoogendoorn, A. F. Frangi, G. Piella, A framework for optimal kernel-based manifold embedding of medical image data, *Computerized Medical Imaging and Graphics* 41 (2015) 93 – 107, *SI: Machine Learning in Medical Imaging*.
- 775 [45] F. R. K. Chung, *Spectral Graph Theory* (CBMS Regional Conference Series in Mathematics, No. 92), American Mathematical Society, 1997.
- [46] A. Shtern, R. Kimmel, Matching the LBO eigenspace of non-rigid shapes via high order statistics, *Axioms* 3 (3) (2014) 300–319.
- 780 [47] D. Mateus, R. Horaud, D. Knossow, F. Cuzzolin, E. Boyer, Articulated shape matching using laplacian eigenfunctions and unsupervised point registration, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
- [48] J. Cardoso, A. Souloumiac, Jacobi angles for simultaneous diagonalization, *SIAM Journal on Matrix Analysis and Applications* 17 (1) (1996) 161–164,
- 785 <http://bsp.teithe.gr/members/downloads/JointDiagonalization.html> (last accessed: 14.01.2016).
- [49] B. Afsari, P. Krishnaprasad, Some gradient based joint diagonalization methods for ICA, in: *Independent Component*

Analysis and Blind Signal Separation, Vol. 3195 of Lecture Notes in Computer Science, 2004, pp. 437–444.

- [50] A. Ziehe, P. Laskov, G. Nolte, K.-R. Müller, A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation, *Journal of Machine Learning Research* 5 (2004) 777–800.
- 790 [51] A. Yeredor, Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation, *IEEE Transactions on Signal Processing* 50 (7) (2002) 1545–1553.
- [52] R. Vollgraf, K. Obermayer, Quadratic optimization for simultaneous matrix diagonalization, *IEEE Transactions on Signal Processing* 54 (9) (2006) 3270–3278.
- [53] M. Zhong, M. Girolami, A Bayesian approach to approximate joint diagonalization of square matrices, arXiv:1206.4666, 795 2012.
- [54] A. Kovnatsky, M. M. Bronstein, A. M. Bronstein, K. Glashoff, R. Kimmel, Coupled quasi-harmonic bases, *Computer Graphics Forum* 32 (24) (2013) 439–448.
- [55] R. A. Horn, C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.
- [56] K. Glashoff, M. M. Bronstein, Matrix commutators: their asymptotic metric properties and relation to approximate joint 800 diagonalization, *Linear Algebra and its Application* 439 (2013) 2503–2513.
- [57] H. Lin, Almost commuting selfadjoint matrices and applications, *Fields Institute Commun.* 13 (1997) 193233.
- [58] N. Sochen, R. Kimmel, R. Malladi, A general framework for low level vision, *IEEE Transactions on Image Processing* 7 (3) (1998) 310–318.
- [59] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: *Int. Conf. on Computer Vision*, 1998, pp. 839–846. 805
- [60] J. J. Koenderink, A. J. V. Doorn, The structure of locally orderless images, *International Journal of Computer Vision* 31 (2) (1999) 159–168.
- [61] S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, J. P. W. Pluim, Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information, *Medical Physics* 35 (4) (2008) 1407–1417.
- 810 [62] T. Sebastian, P. Klein, B. Kimia, Recognition of shapes by editing their shock graphs, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (5) (2004) 550–571, <http://www.lems.brown.edu/vision/researchAreas/SIID/> (last accessed: 02.04.2015).
- [63] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE* 86 (11) (1998) 2278–2324, 815 <http://yann.lecun.com/exdb/mnist/> (last accessed: 06.03.2015).
- [64] M. Mellor, M. Brady, Phase mutual information as a similarity measure for registration, *Medical Image Analysis* 9 (4) (2005) 330–343.
- [65] G.-A. Bilodeau, A. Torabi, P.-L. St-Charles, D. Riahi, Thermalvisible registration of human silhouettes: A similarity measure performance evaluation, *Infrared Physics & Technology* 2013 64 (2014) 79 – 86, 820 <http://vcipl-okstate.org/pbvs/bench/index.html> (last accessed: 18.06.2016).
- [66] B. Menze, et al., The multimodal brain tumor image segmentation benchmark (BRATS), *IEEE Transactions on Medical Imaging* 34 (10) (2015) 1993–2024, <http://www.imm.dtu.dk/projects/BRATS2012/> (last accessed: 25.01.2016).
- [67] L. Mercier, R. F. Del Maestro, K. Petrecca, D. Araujo, C. Haegelen, D. L. Collins, Online database of clinical MR and 825 ultrasound images of brain tumors, *Medical Physics* 39 (6) (2012) 3253–3261.
- [68] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, J. C. Gee, N4itk: Improved N3 bias correction, *IEEE Transactions on Medical Imaging* 29 (6) (2010) 1310–1320.
- [69] H. Rivaz, Z. Karimaghloo, V. S. Fonov, D. L. Collins, Nonrigid registration of ultrasound and MRI using contextual conditioned mutual information, *IEEE Transactions on Medical Imaging* 33 (3) (2014) 708–725.

- 830 [70] W. Wein, A. Ladikos, B. Fuerst, A. Shah, K. Sharma, N. Navab, Global registration of ultrasound to MRI using the LC2 metric for enabling neurosurgical guidance, in: *Medical Image Computing and Computer-Assisted Intervention*, Vol. 8149 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2013, pp. 34–41.
- [71] H. Rivaz, D. L. Collins, Self-similarity weighted mutual information: A new nonrigid image registration metric, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 91–98.
- 835 [72] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, W. Eubank, PET-CT image registration in the chest using free-form deformations, *IEEE Transactions on Medical Imaging* 22 (1) (2003) 120–128.
- [73] A. Gholipour, N. Kehtarnavaz, Computationally efficient mutual information estimation for non-rigid image registration, in: *IEEE Int. Conf. on Image Processing*, 2008, pp. 1792–1795.
- 840 [74] Z. Yi, S. Soatto, Nonrigid registration combining global and local statistics, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2009, pp. 2200–2207.
- [75] M. E. Vandenberghe, A.-S. Hérard, N. Souedet, E. Sadouni, M. D. Santin, D. Briet, D. Carré, J. Schulz, P. Hantraye, P.-E. Chabrier, et al., High-throughput 3d whole-brain quantitative histopathology in rodents, *Scientific reports* 6.
- [76] M. Blesa Cabez, A. Serag, A. G. Wilkinson, D. Anblagan, E. J. Telford, R. Pataky, S. S. Sparrow, G. Macnaught, S. I. Semple, M. E. Bastin, J. P. Boardman, Parcellation of the healthy neonatal brain into 107 regions using atlas propagation through intermediate time points in childhood, *Frontiers in Neuroscience* 10 (220).
- 845 [77] M. Kuklisova-Murgasova, A. Cifor, R. Napolitano, A. Papageorghiou, G. Quaghebeur, M. A. Rutherford, J. V. Hajnal, J. A. Noble, J. A. Schnabel, Registration of 3d fetal neurosonography and MRI, *Medical Image Analysis* 17 (8) (2013) 1137 – 1150.
- 850 [78] S. Jae Hwang, N. Adluru, M. D. Collins, S. N. Ravi, B. B. Bendlin, S. C. Johnson, V. Singh, Coupled harmonic bases for longitudinal characterization of brain networks, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2016.