A Robust Similarity Measure for Volumetric Image Registration with Outliers

Patrick Snape[a,*], Stefan Pszczolkowski[c], Stefanos Zafeiriou[a], Georgios Tzimiropoulos[b], Christian Ledig[a], Daniel Rueckert[a]

[a]*Imperial College London, Department of Computing, London, SW7 2AZ, UK*
[b]*University of Nottingham, School of Computer Science, Nottingham, NG8 1BB, UK*
[c]*University of Nottingham, School of Medicine, Nottingham, NG7 2UH, UK*

Abstract

Image registration under challenging realistic conditions is a very important area of research. In this paper, we focus on algorithms that seek to densely align two volumetric images according to a global similarity measure. Despite intensive research in this area, there is still a need for similarity measures that are robust to outliers common to many different types of images. For example, medical image data is often corrupted by intensity inhomogeneities and may contain outliers in the form of pathologies. In this paper we propose a global similarity measure that is robust to both intensity inhomogeneities and outliers without requiring prior knowledge of the type of outliers. We combine the normalised gradients of images with the cosine function and show that it is theoretically robust against a very general class of outliers. Experimentally, we verify the robustness of our measures within two distinct algorithms. Firstly, we embed our similarity measures within a proof-of-concept extension of the Lucas-Kanade algorithm for volumetric data. Finally, we embed our measures within a popular non-rigid alignment framework based on free-form deformations and show it to be robust against both simulated tumours and intensity inhomogeneities.

Keywords: Image Registration, Lucas-Kanade, Normalised Gradient, Free-form Deformation

1. Introduction

Image registration is an important procedure in many areas of computer vision for both 2D and volumetric 3D images. Given its relevance, there is a large body of prior work concentrating on methodologies for performing accurate registration. In this paper, we are most interested in techniques that attempt to densely align two images according to a global similarity measure. Many global similarity measures have been proposed, yet only a few focus on being robust to the presence of outliers and systematic errors. We consider a similarity measure to be robust if it is not biased by the presence of noise and/or occlusions within the image to be registered. Generally, if a similarity measure is biased by noise or occlusions, this will manifest as a failure to accurately register the images.

In the case of 2D images, systematic errors and outliers are common in the form of illumination variance and occlusions, which naturally occur in so called "in-the-wild" images. Although 2D image alignment is a broad field, much of the existing work focuses on augmenting existing efficient algorithms with improved robust properties.

For example, one of the first algorithms to describe a 2D image alignment approach was the Lucas-Kanade (LK) algorithm [1]. The LK algorithm concentrates on recovering a warp that best maximises a similarity measure between two images. Numerous extensions to the LK algorithm have been proposed [2, 3, 4] and most are based on ℓ_2 norm minimisation [3, 5, 6, 7]. Most notably, the inverse compositional framework proposed by Baker and Matthews [5] provides a computationally efficient framework for solving the least squares problem.

For volumetric, or 3D images, outliers can occur in the form of pathologies, and systematic errors are commonly seen as intensity inhomogeneities caused by image acquisition artefacts such as bias fields [8]. Several methods have been proposed for registration of medical images with mismatches, focusing on robustness [9], tumour models [10] or Bayesian models [11]. However, previous methods [12] all require prior knowledge of what constitutes a mismatch in order to detect and ignore them. Additionally, a number of methods based on mutual information have been proposed to reduce the effect of intensity inhomogeneities in the registration [13, 14, 15].

To the best of our knowledge, no existing similarity measure provides robustness against both outliers and intensity inhomogeneities for registration of 3D images *without prior knowledge of the type of dissimilarity*. However, the 2D similarity measure recently proposed in [16], has been shown to be robust against both general occlusions

---

*Principal corresponding author
Email addresses: `p.snape@imperial.ac.uk` (Patrick Snape), `stefan.pszczolkowskiparraguez@nottingham.ac.uk` (Stefan Pszczolkowski), `s.zafeiriou@imperial.ac.uk` (Stefanos Zafeiriou), `yorgos.tzimiropoulos@nottingham.ac.uk` (Georgios Tzimiropoulos), `christian.ledig@imperial.ac.uk` (Christian Ledig), `d.rueckert@imperial.ac.uk` (Daniel Rueckert)

and illumination variance. This measure is formulated as the cosine of normalised gradient orientations and is simple and efficient to compute. For 3D images, we seek to provide a similarity measure that can utilise the robust properties of the cosine function. This requires calculating a similarity between the two images being aligned that can be represented as an angle. In this work, we provide two separate methodologies of measuring angular similarity between 3D images.

We clarify that when we state 3D images, we are referring to volumetric data where it is valid to compute gradients along all 3 of the principle axes. This may be in contrast to other usages of the term "3D" when referring to data sources such as range images or mesh data. In this case, we would refer to these data sources as 2.5D data, as the computation of the gradient for the depth axis is non-trivial. In fact, as mentioned by Baker et al. [17], the inverse compositional algorithm is no longer valid when extended to 2.5D data due to the representation of the data as a surface. Therefore, given that the treatment of 2.5D data is totally different from the 3D volumetric images that we use here, we do not further consider it.

It is important to note that there is a large amount of existing work on using gradient information for image registration on volumetric data. The concept of normalised image gradients was introduced to the field of medical image registration by Pluim et al. [18]. In [18], normalised mutual information (NMI) [19] is weighted voxelwise by the normalised image gradients in order to incorporate spatial information. After this initial work, the first similarity based solely on normalised gradients was proposed by Haber and Modersitzki [20]. This similarity measure is based on the *squared cosine* of the normalised gradients and is equivalent to minimising the *squared inner product*. In contrast, our proposal is to use the *cosine* of the normalised gradient orientations and is equivalent to minimising the *inner product*. This seemingly small difference, the squaring of the cosine, causes outliers to bias the similarity measure and thus affects the robustness. Despite these properties of the squared cosine measure, it has been successfully utilised in the literature [21, 22, 23] for registering images that do not contain outliers.

Preliminary work on the cosine of orientations has been shown in our previous work [24, 25]. In [24], we gave preliminary results that show that the cosine of normalised gradient orientations represents a robust similarity measure in the presence of both occlusions and intensity inhomogeneities. We extend this work in two major areas:

Firstly, we note that that there are two separate angular measures that can be defined in order to compute the cosine of normalised gradient orientations between two images. These orientations are based on the spherical coordinates of the gradients and the inner product between the gradients. As a proof of concept, we directly extend the methodology of [16] to provide evidence that our similarity measures are robust to occlusions and intensity inhomogeneities. Although the extension of LK-type algorithms
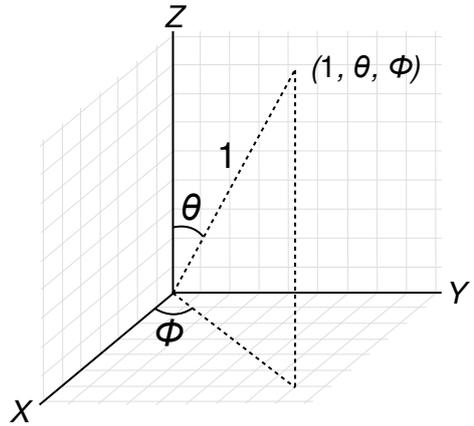
Figure 1: An illustration of the spherical coordinate system as described in this paper.

to 3D is simple and was proposed in [17], no previous investigation has been done on how similarity measures perform when extended to 3D. Since no previous work has investigated the use of 3D Lucas-Kanade for robust registration, we chose to extend existing robust 2D methods into 3D. These extended state-of-the-art 2D methods are then compared against using a synthetic dataset.

Secondly, in order to show that our similarity measures compete with state-of-the-art techniques, we embed them within a widely and successfully used non-rigid registration framework based on free-form deformations (FFD) [26]. This FFD registration framework differs from our LK example as it is a local deformation model and thus contains many thousands of parameters. This large parameter space makes Gauss-Newton optimisation infeasible due to the memory requirements of inverting the Hessian matrix. For this reason, registration techniques that focus on local deformations are generally solved using gradient descent algorithms that incorporate line searches [26]. We compare against the NMI measure [19], DRAMMS [27] and the cosine squared measure [20] and confirm the robustness of the proposed similarity measures on simulated pathological data from a tumour database. **Secondly**, we provide more extensive evaluation into the robust properties of the cosine of normalised gradient orientations when applied to volumetric data.

The remainder of the paper is organised as follows. In Section 2 we discuss relevant related works and consider the advantages of our proposed measures. In Section 3 we give a thorough explanation of the theory behind our work and empirically verify its robustness on our chosen data. In Section 4 we describe the application of our similarity measure within the Lucas-Kanade algorithm and in Section 5 we show its application within the non-rigid framework of Rueckert et al. [26]. Experimental results within both rigid and non-rigid alignment are given in Section 6. Finally, we draw conclusions in Section 7.

2. Related Work

In this paper we are most interested in similarity measures that have been proposed for parametric image alignment. This is primarily due to the fact that parametric image alignment methods are often very computationally efficient. In particular, we are interested in approaches that show robustness to both intensity inhomogeneities and occlusions.

Within 2D image registration there has been a large body of work that focuses on illumination invariance. Within the most popular parametric image alignment framework, the Lucas-Kanade algorithm (LK), one of the the earliest attempts was by Hager and Belhumeur [28]. In [28], the authors incorporate a linear appearance basis that models illumination variance and excelled in the area of object tracking. However, this requires prior knowledge of the type of object under consideration so that the appearance basis can be precomputed. There is a large amount of existing literature about incorporating prior knowledge via linear bases within the LK algorithm. However, we do not consider them here, as we are most interested in modelling unseen structured variation via robust similarity measures and thus prior knowledge is generally not available.

Within the LK framework, a number of robust measures have been proposed. Black and Jepson [29] proposed incorporating robust statistics into the LK framework and showed their effectiveness in the presence of illumination variance. Dowson and Bowden [3] incorporated the mutual information (MI) measure within the LK framework and found it to be superior to an SSD based measure for illumination variance. However, using mutual information as a similarity measure requires updating the Jacobian and Hessian matrices at each iteration. For this reason, the MI measure is unable to make use of more computationally efficient LK algorithms such as the inverse compositional (IC) method proposed in [5]. Although the authors do give an ad hoc solution that involves fixing the values of the Jacobian and Hessian matrices, it is still a more complex implementation than the original inverse compositional algorithm. Evangelidis and Psarakis [6] provide a correlation measure between images and a computational framework that is invariant to illumination differences. This is a great strength of the algorithm, however, the correlation measure is still grossly affected by outliers. Lucey et al. [4] propose a method of weighting the LK fitting with a large bank of filters in a computationally efficient manner. This is a very effective technique and is largely illumination invariant, however, it gives little benefit when presented with occlusions.

For 3D images, robust registration is a very active area of research. The most commonly used techniques for multi-modal registration are based on mutual information (MI) [30]. These techniques are highly effective at registration of objects with inherent structure such as anatomy but are very sensitive to global corruption such as intensity inhomogeneities. To overcome this, gradient information is often utilised and in particular was incorporated into a MI framework by Pluim et al. [18]. In particular, gradient information helps capture the local structure within an image which is not described by general MI-based registration techniques. Gradient information has also been successfully used in a number of other works [31, 32, 33]. However, these works focus on capturing local structure and not on robustness to artefacts such as pathologies caused by tumours. The most related work is that of Haber and Modersitzki [20], which proposes a similarity measure based on the square of the cosine (inner product squared). We conduct a thorough comparison with this technique and show that it is biased by the presence of occlusions.

Finally, the work of Tzimiropoulos et al. [16] introduced the first similarity measure based on the cosine of normalised gradients. We would stress that although our work is inspired by [16], the calculation of the orientations for our proposed similarities is very different. In particular, it is important to note that calculating an orientation in 3D is more complex than the 2D case due to the extra degree of freedom. In this work, we give a detailed explanation of how to calculate these orientations in 3D and how to optimise them for use in image registration.

3. Cosine of Normalised Gradients

In this section, we describe the concept of the cosine of normalised gradients and specify how they represent a robust measure of similarity. In this work, we consider a similarity measure to be robust if it suppresses the contribution of comparisons between image areas that are unrelated. More specifically, we seek a measure that, when given two images that are visually dissimilar, will calculate zero correlation between them. For example, consider Figure 3 which shows cross sections of a brain containing a tumour. When registering this corrupted image with an image of a healthy brain, the ideal registration would not be biased by the presence of the tumour, as it does not share relevant anatomical structures with the healthy brain. To this end, Tzimiropoulos et al. [34, 16] proposed the cosine of orientation differences between two images, which we describe in detail below.

3.1. Cosine Similarity in 2D

Assuming we are given two 2D images, denoted as I_i $i \in \{1, 2\}$, we define $G_{i,x} = \mathbf{F}_x * I_i$ and $G_{i,y} = \mathbf{F}_y * I_i$ as the gradients obtained by convolving I_i with differentiation approximation filters $\mathbf{F}_x$ and $\mathbf{F}_y$ respectively. We denote the lexicographical vectorisation of $G_{i,x}$ as $\mathbf{g}_{i,x}$ and define an index k into the vector, $\mathbf{g}_{i,x}(k)$. We define an identical vector for $G_{i,y}$ as $\mathbf{g}_{i,y}$. We also define $\mathbf{g}_i(k)$ as the vector formed by concatenating the x and y gradients together. Trivially, we can define the normalised gradient as $\tilde{\mathbf{g}}_i(k) = \frac{\mathbf{g}_i(k)}{\|\mathbf{g}_i(k)\|}$ where $\|\mathbf{g}_i(k)\| = \sqrt{g_{i,x}(k)^2 + g_{i,y}(k)^2}$. We
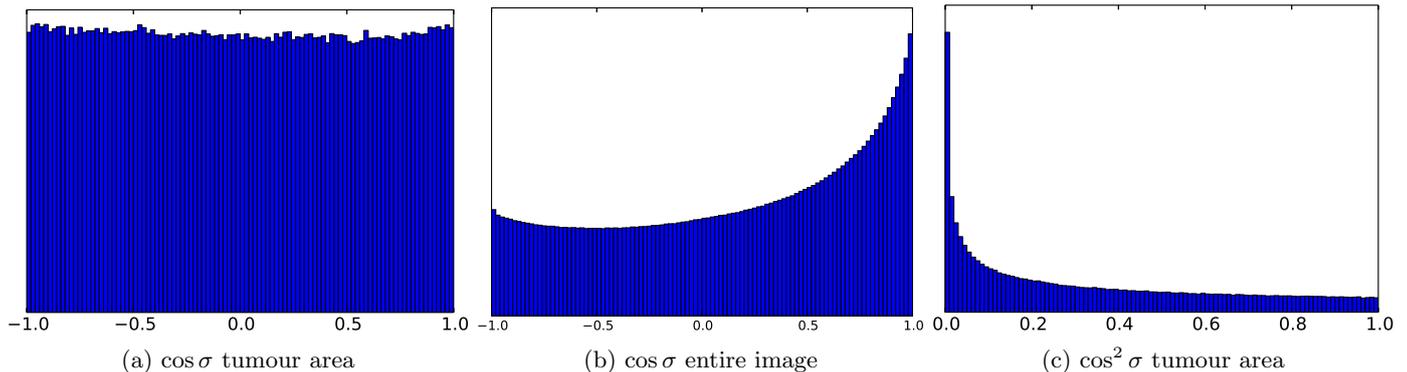
Figure 2: The distributions of $\cos\sigma$ and $\cos^2\sigma$ averaged over 10 subjects from the BraTS simulated images. The images were registered using a rigid transformation prior to computation and only the tumour areas were sampled. (a) shows the distribution of $\cos\sigma$ in the simulated tumour region. (b) shows the distribution of $\cos\sigma$ over the entire image. (c) shows the distribution of $\cos^2\sigma$ proposed in [20] in the simulated tumour region

also define similar vectors for the x and y components separately, with $\tilde{\mathbf{g}}_{i,x}$ being the x components concatenated in lexicographical ordering and $\tilde{\mathbf{g}}_{i,y}$ being the y components. Finally, $\tilde{\mathbf{g}}_i$ is the vector of concatenated normalised gradients for image I_i.

Given the normalised gradients, it is simple to parametrise them within a polar coordinate system with radius $r_i(k) = \|\tilde{\mathbf{g}}_i(k)\| = 1$, orientation $\phi_i(k) = \arctan\frac{\tilde{g}_{i,y}(k)}{\tilde{g}_{i,x}(k)}$ and pole at the origin. Given orientations from two dissimilar images, it is reasonable to assume that difference between the orientations, $\Delta\phi(k) = \phi_1(k) - \phi_2(k)$, can take any angle between $[0, 2\pi]$. Intuitively, this implies that selecting two pixels from dissimilar images is unlikely to yield any correlation between the images. In [16], it was experimentally verified that the orientation differences follow a uniform distribution, $\Delta\phi(k) \sim U(0, 2\pi)$. The fact that the orientation differences follows a uniform distribution is unsurprising under the assumption that the two images have absolutely no correlation. However, the expectation of the cosine of the uniform distribution is zero, which is a powerful property that can be exploited for image registration. It is powerful because it means that the expected overall contribution of uncorrelated areas to any cost function will be zero, meaning the uncorrelated areas do not affect the result of the registration.

Formally, we assume $\Delta\phi(k)$ is a stationary random process $y(t)$ with index $t \triangleq k \in \mathcal{R}^2$, where $\forall t \sim U(0, 2\pi)$. We define the random process $z(t) = \cos y(t)$ and thus $\forall t$ random variable $Z = z(t)$ has mean value $E\{Z\} = 0$. In fact, by assuming mean ergodicity, we find that

$$E\{Z\} \propto \int z(t)dt \equiv \int_{\mathcal{R}^2} \cos[\Delta\phi(k)]dk = 0 \qquad (1)$$

This is an important property for a similarity measure to be robust against occlusions. Since occlusions do not provide useful information for alignment, ideally they would be ignored. However, manual segmentation of occluded areas is time consuming and prone to error. Therefore, an ideal robust similarity measure would be able to auto-matically identify regions of the image that are occluding the true object of interest. Under the previous definition of robustness, the cosine similarity naturally represents a robust similarity measure as it automatically suppresses the contribution of outliers.

Given an image warping function with parameters $\mathbf{p}$, maximising the sum of the cosine of orientation differences provides the robust similarity measure:

$$q = \sum_k \cos(\Delta\phi(k)[\mathbf{p}]) \qquad (2)$$

For more details of the specifics of optimising (2) for image alignment, we refer the reader to [16].

3.2. Cosine Similarity in 3D

We make very similar assumptions for 3D images as we did in Section 3.1 for 2D images. We simply extend the previous notation by including the gradient of the z-axis, denoted as $G_{i,z} = \mathbf{F}_z * I_i$. We also redefine the normalised gradient as $\tilde{\mathbf{g}}_i(k) = \frac{\mathbf{g}_i(k)}{\|\mathbf{g}_i(k)\|}$ where $\|\mathbf{g}_i(k)\| = \sqrt{g_{i,x}(k)^2 + g_{i,y}(k)^2 + g_{i,z}(k)^2}$ and $\mathbf{g}_i(k)$ is defined as the vector formed by concatenating the x, y and z gradients together.

Measuring the angular distance between vectors in 3D is more complex than in 2D, due to the extra degree of freedom. In the following sections, we describe two different measures that can be used to calculate similarities between vectors within 3D images, the spherical coordinates and the inner product. In the previous section, we described in detail how properties of the cosine of a uniform distribution can be exploited to form a robust measure of similarity. The most important property was that uncorrelated areas such as occlusions should have no impact registration. This was formalised as the expectation of the sum of the uncorrelated elements should be zero. In the case of input to the cosine function, a given distribution must simply be symmetric over the positive and negative span of outputs of the cosine. When symmetric over the

4

positive and negative outputs, the expectation of the cosine function is zero. In fact, we can relax the definition of a measure being robust to outliers by stating that we desire a measure whereby the expectation of the measure over image areas that are uncorrelated is zero.

In practise, when comparing two images where one image contains occlusions, there will be regions that are correlated and then the occluded region that is uncorrelated. In this case, the total distribution of all pixels will be described by a mixture model of the occluded and non-occluded regions. We desire that the distribution of the uncorrelated areas has an expectation of zero and thus will not affect the optimisation of the similarity measure.

In Section 3.2.1 and Section 3.2.2 we describe two measures of angular difference between 3D images. We investigate the distribution of these angular measures when combined with the cosine function and motivate that they are both suitable for use as a similarity measure between real 3D images.

3.2.1. Spherical Coordinates

In 2D, a natural parametrisation of the angle between the two gradient vectors is the polar coordinate system. In 3D, we have three gradient vectors and thus require two angles to describe their orientation. Unlike in 2D, where the vectors lie on the unit circle, in 3D the vectors lie on the surface of a unit sphere. Therefore, it is possible to parametrise the vectors in terms of the spherical coordinate system, which is described by two angles: the azimuth angle ϕ with range $[0, 2\pi]$ and the elevation angle θ with range $[0, \pi]$. Given the normalised gradients as vectors with Cartesian coordinates, we can calculate the spherical angles as follows:

$$
\begin{aligned}
r_i(k) &= \|\tilde{\mathbf{g}}_i(k)\| = 1 \\
\phi_i(k) &= \arctan \frac{\tilde{g}_{i,y}(k)}{\tilde{g}_{i,x}(k)} \\
\theta_i(k) &= \arccos \tilde{g}_{i,z}(k)
\end{aligned}
\qquad (3)
$$

An illustration of the spherical coordinate system, as used in this paper, is given in Figure 1.

Our proposal is to combine the spherical coordinates with the cosine function in order to provide a robust similarity measure. Similar to the 2D case,, we propose the cosine of azimuth differences, $\Delta\phi = \phi_1 - \phi_2$, and the cosine of elevation differences, $\Delta\theta = \theta_1 - \theta_2$, as a combined similarity measures. Given a 3D image warping function with parameters $\mathbf{p}$, the spherical coordinates form a similarity measure as follows:

$$
q = \sum_k \cos\left(\Delta\phi(k)[\mathbf{p}]\right) + \sum_k \cos\left(\Delta\theta(k)[\mathbf{p}]\right) \qquad (4)
$$

Optimisation of (2) is described in detail in Section 4.

Experimentally, we verified that $\Delta\phi$ approximates a symmetric distribution for simulated tumour data taken from the Multimodal Brain Tumor Image Segmentation
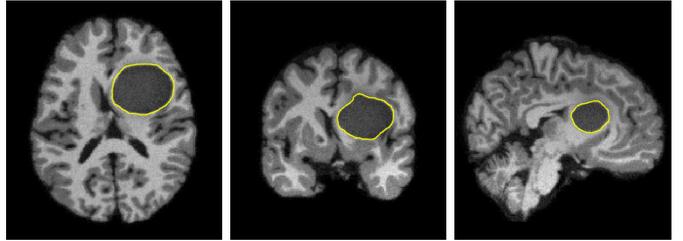
Figure 3: Example images of a T1-weighted brain containing a tumour area. The tumour areas are outlined in yellow in each image. Left: Axial View. Middle: Coronal View. Right: Sagital View.

(BraTS) challenge, as shown in Figure 4. Figure 4a shows the distribution of $\Delta\phi$ between the tumour area circled in yellow in Figure 3 and a healthy brain. The images were registered using a rigid transformation before $\Delta\phi$ was computed. The azimuth angle is analogous to the angle studied in [34] and follows the same uniform distribution, $\Delta\phi \sim U(0, 2\pi)$.

When the entire region of the rigidly registered brain images is considered, we find that the distribution of $\Delta\phi$ is clearly a mixture of two separate models, one for the occluded area and one for the rigidly registered area. Figure 4b shows the distribution of $\Delta\phi$ calculated over the entire image region of each image and a Laplacian distribution that best fits the data. Thus, our experimental evidence suggests that the total distribution of $\Delta\phi$ over the entire image region is a mixture model between a uniform distribution and a Laplacian distribution with approximately zero mean.

3.2.2. Inner Product

A more general angular measure between two vectors is the inner product. Unlike in [16] or Section 3.2.1, the inner product is a single angle and not the difference between two angles. Practically, the inner product measures the projection error between two vectors and is defined as:

$$
\cos \sigma = \tilde{\mathbf{g}}_1^\top \tilde{\mathbf{g}}_2 \qquad (5)
$$

In [34], the authors reasonably propose that the angle between the gradients of dissimilar images can take any value in $[0, 2\pi)$ with equal probability. Similarly, the relationship between the gradient vectors of two dissimilar 3D images could feasibly be in any direction with equal probability. Therefore, the distribution of inner products between two unrelated vectors can take the values $[-1, 1]$ with equal probability. Due to the expected range of inner product values, we would expect that $\cos \sigma$ follows a uniform distribution, $\cos \sigma \sim U(-1, 1)$. Note that this is a different assumption to that made in [34], which assumes that the *azimuth angle itself*, $\Delta\phi$, follows a uniform distribution. However, it is merely sufficient that the total sum of values from the dissimilar vectors is zero. Therefore, since $E\{U(-1, 1)\} = 0$, the inner product of normalised gradients satisfies our definition of being robust to outliers.

5
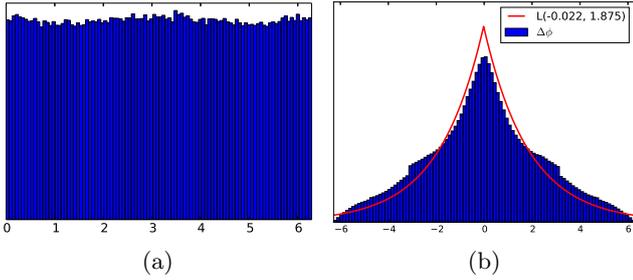
Figure 4: The mean distribution of $\Delta\phi$ of the BraTS simulated images. The images were registered using a rigid transformation and only the tumour areas were sampled. (a) the distribution of $\Delta\phi$ in the simulated tumour region. (b) the distribution of $\Delta\phi$ over the entire image. It also shows the Laplacian distribution that best fits the data.

In Figure 2a, we show that this assumption holds for the simulated tumour data taken from the BraTS challenge.

When the entire region of the rigidly registered brain images is considered, we find that the distribution of $\cos\sigma$ is clearly a mixture of two separate models, one for the occluded area and one for the rigidly registered area. Figure 2b shows the distribution of $\cos\sigma$ calculated over the entire image region. In this case, the distribution of the inner product appears to be a mixture model between a uniform distribution and a zero mean Laplacian distribution. However, due to the ambiguity in the inner product in terms of orientation, the angle of the inner product is only defined in the range $[0, \pi]$ and thus only the positive tail of the Laplacian appears.

In Figure 2c we also show the distribution of the similarity measure proposed by Haber and Modersitzki [20]. In [20], the authors propose the inner product as a similarity measure, which looks very similar to the measure we proposed in Equation (5). However, Haber and Modersitzki [20] maximise the square of the inner product using a least squares Gauss-Newton optimisation. As we have shown, the inner product is related to the cosine between the vectors. Haber and Modersitzki [20] proposed the inner product squared as a similarity, which is equivalent to the *square of the cosine*. As we can see in Figure 2c, the cosine squared does not represent a symmetric distribution and therefore is not a robust similarity measure by our definition.

4. Robust Lucas-Kanade

Little work has been published on the applications of 3D Lucas-Kanade (LK), despite the extension of LK into 3D being trivial [17]. Given the relative efficiency of Lucas-Kanade and other similar algorithms, and their potential accuracy under challenging conditions, we propose to investigate using LK for robust affine alignment in voxel images. To the best of our knowledge, this is the first such comparison of using state-of-the-art LK fitting algorithms for voxel data. Given the similarity measures defined in

Section 3, we propose novel LK algorithms that directly maximise the measures.

When referring to the operations performed by the LK algorithm we will use the following notations. Warp functions $\mathcal{W}(\mathbf{x}_i; \mathbf{p}) = [\mathcal{W}_x(\mathbf{x}_i; \mathbf{p}), \mathcal{W}_y(\mathbf{x}_i; \mathbf{p}), \mathcal{W}_z(\mathbf{x}_i; \mathbf{p})]$ express the warping of the ith 3D coordinate vector $\mathbf{x}_i = [x_i, y_i, z_i]^\top$ by a set of parameters $\mathbf{p} = [p_1, \ldots, p_n]^\top$, where n is the number of warp parameters. We extend the previously defined linear index k in to a coordinate vector, $\mathbf{x} = [x_1, y_1, z_1, \ldots, x_D, y_D, z_D]$ that represents the concatenated vector of coordinates, of length D, which allows the definition of a single warp for an entire image, $\mathcal{W}(\mathbf{x}; \mathbf{p}) = [\mathcal{W}_x(\mathbf{x}_1; \mathbf{p}), \mathcal{W}_y(\mathbf{x}_1; \mathbf{p}), \mathcal{W}_z(\mathbf{x}_1; \mathbf{p}),$ $\ldots, \mathcal{W}_x(\mathbf{x}_D; \mathbf{p}), \mathcal{W}_y(\mathbf{x}_D; \mathbf{p}), \mathcal{W}_z(\mathbf{x}_D; \mathbf{p})]$. We assume that the identity warp is found when $\mathbf{p} = \mathbf{0}$, which implies that $\mathcal{W}(\mathbf{x}; \mathbf{0}) = \mathbf{x}$. We abuse notation and define the warping of an image I by parameter vector $\mathbf{p}$ as $I(\mathbf{p}) = I(\mathcal{W}(\mathbf{x}; \mathbf{p}))$, where $I(\mathbf{p})$ is a single column vector of concatenated pixels.

In the following sections we describe the details of relevant LK algorithms. We begin with the original forward additive LK algorithm [1, 5]. We then describe the enhanced correlation coefficient (ECC) algorithm [6] which was shown to be robust to intensity inhomogeneities. We conclude the existing algorithms with a description of the efficient inverse compositional algorithms for both the original LK method [35] and the ECC method [6]. Following the existing literature, we present our proposed LK algorithms. The first of which is a variation of the ECC method for normalised gradients and the second involves a traditional inverse compositional method. Both of the proposed algorithms take the inverse compositional form due to its computational efficiency.

4.1. Forward Additive LK Fitting

The original forward additive ℓ_2 LK algorithm [1, 5] seeks to minimise the sum of squared differences (SSD) between a given template image and an input image by minimising the sum of the squared pixel differences:

$$\underset{\mathbf{p}}{\arg\min} \quad \|I(\mathbf{p}) - T(\mathbf{0})\|^2 \tag{6}$$

where $T(\mathbf{0})$ is the unwarped reference template image. Due to the non-linear nature of (6) with respect to $\mathbf{p}$, (6) is linearised by taking the first order Taylor series expansion. By iteratively solving for some small $\Delta\mathbf{p}$ update to $\mathbf{p}$, the objective function becomes

$$\underset{\mathbf{p}}{\arg\min} \quad \|I(\mathbf{p}) + \nabla I \frac{\partial \mathcal{W}}{\partial \mathbf{p}} \Delta\mathbf{p} - T(\mathbf{0})\|^2 \tag{7}$$

where ∇I is the gradient over each dimension of $I(\mathbf{p})$ warped into the frame of T by the current warp estimate $\mathcal{W}(\mathbf{x}; \mathbf{p})$. $\frac{\partial \mathcal{W}}{\partial \mathbf{p}}$ is the Jacobian of the warp and represents the first order partial derivatives of the warp with respect to each parameter. $\nabla I \frac{\partial \mathcal{W}}{\partial \mathbf{p}}$ is commonly referred to as the

steepest descent images. We will express the steepest descent images as $\frac{\partial I(\mathbf{p})}{\partial \mathbf{p}}$. Equation (7) is now solvable by assuming the Gauss-Newton approximation to the Hessian, $\mathbf{H} = \left[\frac{\partial I(\mathbf{p})}{\partial \mathbf{p}}^\top \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}} \right]$:

$$\Delta \mathbf{p} = \mathbf{H}^{-1} \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}}^\top [T(\mathbf{0}) - I(\mathbf{p})] \tag{8}$$

Equation (8) can then be solved by iteratively updating $\mathbf{p} \leftarrow \mathbf{p} + \Delta \mathbf{p}$ until convergence.

4.2. ECC LK Fitting

The enhanced correlation coefficient (ECC) measure, proposed by Evangelidis and Psarakis [6], seeks to be invariant to illumination differences between the input and template image. This is done by suppressing the magnitude of each pixel through normalisation. In [6], they provide the following cost function

$$\underset{\mathbf{p}}{\operatorname{argmax}} \quad \frac{I(\mathbf{p})^\top T(\mathbf{0})}{\|I(\mathbf{p})\| \|T(\mathbf{0})\|} \tag{9}$$

Assuming a delta update as before and linearising in a similar manner to (7) results in

$$\underset{\mathbf{p}}{\operatorname{argmax}} \quad \hat{T} \frac{I(\mathbf{p}) + \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}} \Delta \mathbf{p}}{\|I(\mathbf{p}) + \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}} \Delta \mathbf{p}\|} \tag{10}$$

where $\hat{T} = \frac{T(\mathbf{0})}{\|T(\mathbf{0})\|}$. Evangelidis and Psarakis [6] give a very comprehensive proof of the upper bound of Equation (10), which yields the following solution for $\Delta \mathbf{p}$

$$\Delta \mathbf{p} = \mathbf{H}^{-1} \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}}^\top \left[\frac{\|I(\mathbf{p})\|^2 - I(\mathbf{p})^\top \mathbf{Q} I(\mathbf{p})}{\hat{T}^\top I(\mathbf{p}) - \hat{T}^\top \mathbf{Q} I(\mathbf{p})} \hat{T} - I(\mathbf{p}) \right] \tag{11}$$

where $\mathbf{Q}$ is an orthogonal projection operator on the Jacobian, $\mathbf{J} = \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}}$, defined as $\mathbf{Q} = \mathbf{J}(\mathbf{J}^\top \mathbf{J})^{-1} \mathbf{J}^\top$.

In fact, the Δp update given in [6] is more complex than (11), as it seeks to find an upper bound on the correlation between the two images. However, in the case where (11) does not apply, it is unlikely that the algorithm is able to converge. For this reason, we only consider the update equation presented in (11).

4.3. Inverse Compositional LK

The inverse compositional algorithm, proposed by Baker and Matthews [35], performs a compositional update of the warp and linearises over the template rather than the input image. Linearisation of the template image causes the gradient in the steepest descent images term to become fixed. The compositional update of the warp assumes linearisation of the term $\frac{\partial \mathcal{W}(\mathbf{x};\mathbf{0})}{\partial \mathbf{p}}$, which is also fixed. Therefore, the entire Jacobian term, and by extension the Hessian matrix, are also fixed. Similar to the ℓ_2 SSD algorithm

described in Section 4.1, we pose the objective function as:

$$\underset{\mathbf{p}}{\operatorname{argmin}} \|T(\Delta \mathbf{p}) - I(\mathbf{p})\|^2 \tag{12}$$

where we notice that the roles of the template and input image have been swapped. Assuming an inverse compositional update to the warp, $\mathcal{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathcal{W}(\mathbf{x};\mathbf{p}) \circ \mathcal{W}(\mathbf{x};\Delta \mathbf{p})^{-1}$ and linearisation around the template, (12) can be expanded as:

$$\underset{\mathbf{p}}{\operatorname{argmin}} \quad \|I(\mathbf{p}) - \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}} \Delta \mathbf{p} - T(\mathbf{0})\|^2 \tag{13}$$

Solving for $\Delta \mathbf{p}$ is identical to (8), except the Jacobian and Hessian have been pre-computed

$$\Delta \mathbf{p} = \mathbf{H}^{-1} \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}}^\top [I(\mathbf{p}) - T(\mathbf{0})] \tag{14}$$

The ECC can also be described as an inverse compositional algorithm, by performing the same update to the warp and simply swapping the roles of the template and reference image. In short, solving ECC in the inverse compositional case becomes

$$\Delta \mathbf{p} = \mathbf{H}^{-1} \frac{\partial \hat{T}}{\partial \mathbf{p}}^\top \left[\frac{\|\hat{T}\|^2 - \hat{T}^\top \mathbf{Q} \hat{T}}{I(\mathbf{p})^\top \hat{T} - I(\mathbf{p})^\top \mathbf{Q} \hat{T}} I(\mathbf{p}) - \hat{T} \right] \tag{15}$$

where $\mathbf{Q}$ is as before, except $\mathbf{J} = \frac{\partial \hat{T}}{\partial \mathbf{p}}$. Any term involving $\hat{T}$ is fixed and pre-computable, so the reduction of calculations per-iteration is substantial.

It is worth noting that not every family of warps is suitable for the inverse compositional approach. The warp must belong to a family that forms a group, and the identity warp must exist in the set of possible warps. For more complex warps, such as piecewise affine and thin plate spline warping, approximations to the inverse compositional updates have been proposed [36, 37].

4.4. Inner Product ECC LK

Given the inner product similarity measure as described in Section 3.2.2, we seek to embed it within the LK framework in order to present a robust parametric alignment algorithm. Therefore, we begin by restating our cost function:

$$\underset{\mathbf{p}}{\operatorname{argmax}} \quad \sum_k \cos\left(\sigma(\mathbf{p})\right) \tag{16}$$

with an abuse of notation for the parameters which are hidden within the cosine function. Expanding (16) reveals the parameters and makes the relationship between our inner product similarity and the ECC framework clear:

$$\underset{\mathbf{p}}{\operatorname{argmax}} \quad \tilde{\mathbf{g}}_I(\mathbf{p})^\top \tilde{\mathbf{g}}_T(\mathbf{0}) \tag{17}$$

which yields our forward additive algorithm.

However, unlike in ECC where the vectors represent concatenated normalised intensities, we are considering

normalised gradients. Since gradients have three separate components we must consider the derivatives when linearising $\mathbf{g}_I$. Since $\mathbf{g}_I$ is composed of multiple components, there will be extra derivatives to calculate via the chain rule. Formally, linearising (17) with respect to $\mathbf{g}_I$ yields

$$\operatorname*{argmax}_{\mathbf{p}} \quad \tilde{\mathbf{g}}_T \frac{\tilde{\mathbf{g}}_I(\mathbf{p}) + \mathbf{J_g}\Delta\mathbf{p}}{\|\tilde{\mathbf{g}}_I(\mathbf{p}) + \mathbf{J_g}\Delta\mathbf{p}\|} \quad (18)$$

where $\mathbf{J_g}$ is the matrix formed by correctly computing the derivative of $\mathbf{g}_I$ with respect to each component of $\mathbf{g}_I$. For example, given that $\mathbf{g}_{I,x}(\mathbf{p})$ is a vector formed of the x-components of the gradients and $\tilde{\mathbf{g}}_{I,x}(\mathbf{p})$ is a vector formed of $\tilde{\mathbf{g}}_{I,x}(\mathbf{p})[k] = \frac{\mathbf{g}_{I,x}(\mathbf{p})[k]}{\|\mathbf{g}_I(\mathbf{p})[k]\|}$, the true derivative of $\tilde{\mathbf{g}}_{I,x}(\mathbf{p})$ is

$$\frac{\partial \tilde{\mathbf{g}}_{I,x}(\mathbf{p})}{\partial \mathbf{p}} = \frac{\partial \tilde{\mathbf{g}}_{I,x}(\mathbf{p})}{\partial \mathbf{g}_{I,x}(\mathbf{p})} \frac{\partial \mathbf{g}_{I,x}(\mathbf{p})}{\partial \mathbf{p}}$$

$$\frac{\partial \tilde{\mathbf{g}}_{I,x}(\mathbf{p})}{\partial \mathbf{g}_{I,x}(\mathbf{p})} = \frac{\mathbf{g}_{I,y}(\mathbf{p})^2 + \mathbf{g}_{I,z}(\mathbf{p})^2}{\left(\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2 + \mathbf{g}_{I,z}(\mathbf{p})^2\right)^{3/2}} \quad (19)$$

where $\frac{\partial \mathbf{g}_{I,x}(\mathbf{p})}{\partial \mathbf{p}}$ is equivalent to $\frac{\partial I(\mathbf{p})}{\partial \mathbf{p}}$ in the original ECC equations. The y and z derivatives are given in a similar fashion as

$$\frac{\partial \tilde{\mathbf{g}}_{I,y}(\mathbf{p})}{\partial \mathbf{g}_{I,y}(\mathbf{p})} = \frac{\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,z}(\mathbf{p})^2}{\left(\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2 + \mathbf{g}_{I,z}(\mathbf{p})^2\right)^{3/2}}$$

$$\frac{\partial \tilde{\mathbf{g}}_{I,z}(\mathbf{p})}{\partial \mathbf{g}_{I,z}(\mathbf{p})} = \frac{\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2}{\left(\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2 + \mathbf{g}_{I,z}(\mathbf{p})^2\right)^{3/2}} \quad (20)$$

However, $\nabla\mathbf{g}_{I,x}$, formally the image gradient, represents the gradient over only the x-component, and is equivalent to the second order derivative of the gradients with respect to x.

Since $\frac{\partial \mathbf{g}_{I,x}(\mathbf{p})}{\partial \mathbf{p}}$ is a matrix and $\frac{\partial \tilde{\mathbf{g}}_{I,x}(\mathbf{p})}{\partial \mathbf{g}_{I,x}(\mathbf{p})}$ is a vector, we multiply the two using a Hadamard product, denoted by the $\odot$ operator. However, $\frac{\partial \tilde{\mathbf{g}}_{I,x}(\mathbf{p})}{\partial \mathbf{g}_{I,x}(\mathbf{p})}$ must first form a matrix, $\mathbf{J}_x$, of size $D \times p$ by repeating the vector p times to form columns within the matrix. Finally the total x-component Jacobian is given by $\mathbf{J}_{\mathbf{g},x} = \mathbf{J}_x \odot \frac{\partial \mathbf{g}_{I,x}(\mathbf{p})}{\partial \mathbf{p}}$.

Given that $\mathbf{J}_{\mathbf{g},i} \forall i \in \{x,y,z\}$ have been calculated, the total derivative term is given by $\mathbf{J_g} = [\mathbf{J}_{\mathbf{g},x}, \mathbf{J}_{\mathbf{g},y}, \mathbf{J}_{\mathbf{g},z}]^\top$. Solving for $\Delta\mathbf{p}$ is now identical to the ECC formulation:

$$\Delta\mathbf{p} = \mathbf{H}^{-1}\mathbf{J_g}\left[\frac{\|\mathbf{g}_I(\mathbf{p})\|^2 - \mathbf{g}_I(\mathbf{p})^\top\mathbf{Q}\mathbf{g}_I(\mathbf{p})}{\tilde{\mathbf{g}}_T^\top\mathbf{g}_I(\mathbf{p}) - \tilde{\mathbf{g}}_T^\top\mathbf{Q}\mathbf{g}_I(\mathbf{p})}\tilde{\mathbf{g}}_T - \mathbf{g}_I(\mathbf{p})\right] \quad (21)$$

Since the update step is identical to the one given in (15) it is simple to reformulate the inner product ECC in an inverse compositional form by following a derivation identical to Section 4.3. In the experimental section, we consider the inverse compositional form of the algorithm.

4.5. Spherical SSD LK

In contrast to the inner product derivation in the previous section, the spherical representation requires the optimisation of the summation of two cosine correlations. In theory, it would be possible to solve for each correlation separately in a manner similar to that proposed by Tzimiropoulos et al. [16]. This would be suboptimal as it would require an alternating optimisation scheme. Therefore, in the interest of solving a single objective, we note the following relationship between the two summations:

$$\operatorname*{argmax}_{\mathbf{p}} \quad \sum_k \cos\left(\Delta\phi(\mathbf{p})\right) + \sum_k \cos\left(\Delta\theta(\mathbf{p})\right) \quad (22)$$

is equivalent to the minimisation of

$$\operatorname*{argmin}_{\mathbf{p}} \left\|\begin{pmatrix}\cos\phi_I(\mathbf{p})\\\sin\phi_I(\mathbf{p})\\\cos\theta_I(\mathbf{p})\\\sin\theta_I(\mathbf{p})\end{pmatrix} - \begin{pmatrix}\cos\phi_T(\mathbf{0})\\\sin\phi_T(\mathbf{0})\\\cos\theta_T(\mathbf{0})\\\sin\theta_T(\mathbf{0})\end{pmatrix}\right\|^2 \triangleq$$

$$\operatorname*{argmin}_{\mathbf{p}} \left\|\begin{pmatrix}\tilde{\mathbf{g}}_{I,x}(\mathbf{p})\\\tilde{\mathbf{g}}_{I,y}(\mathbf{p})\\\tilde{\mathbf{g}}_{I,z}(\mathbf{p})\\\sqrt{1-\tilde{\mathbf{g}}_{I,z}^2(\mathbf{p})}\end{pmatrix} - \begin{pmatrix}\tilde{\mathbf{g}}_{T,x}(\mathbf{0})\\\tilde{\mathbf{g}}_{T,y}(\mathbf{0})\\\tilde{\mathbf{g}}_{T,z}(\mathbf{0})\\\sqrt{1-\tilde{\mathbf{g}}_{T,z}^2(\mathbf{0})}\end{pmatrix}\right\|^2 \quad (23)$$

where $\sin\theta_*(\cdot) = \sqrt{1-\tilde{\mathbf{g}}_{*,z}^2(\cdot)}$ where $*$ denotes either the template or the input image. For notational simplicity, let $\tilde{\mathbf{sz}}(\cdot) = \sqrt{1-\tilde{\mathbf{g}}_z^2(\cdot)}$.

We define the forward additive objective function as

$$\operatorname*{argmin}_{\mathbf{p}}\|\hat{\mathbf{g}}_I(\mathbf{p}) - \hat{\mathbf{g}}_T(\mathbf{0})\|^2 \quad (24)$$

where $\hat{\mathbf{g}}_I(\mathbf{p}) = \left[\tilde{\mathbf{g}}_{I,x}(\mathbf{p}), \tilde{\mathbf{g}}_{I,y}(\mathbf{p}), \tilde{\mathbf{g}}_{I,z}(\mathbf{p}), \tilde{\mathbf{g}}_{I,sz}(\mathbf{p})\right]^\top$ and $\hat{\mathbf{g}}_T(\mathbf{0}) = \left[\tilde{\mathbf{g}}_{T,x}(\mathbf{0}), \tilde{\mathbf{g}}_{T,y}(\mathbf{0}), \tilde{\mathbf{g}}_{T,z}(\mathbf{0}), \tilde{\mathbf{g}}_{T,sz}(\mathbf{0})\right]^\top$, the concatenated vectors of each normalised component.

Similar to the derivation in Section 4.4, the Jacobian must be taken over each component and thus linearising around $\hat{\mathbf{g}}_I(\mathbf{p})$ yields

$$\operatorname*{argmin}_{\mathbf{p}}\|\hat{\mathbf{g}}_I(\mathbf{p}) + \hat{\mathbf{J}}_{\mathbf{g}}\Delta\mathbf{p} - \hat{\mathbf{g}}_T(\mathbf{0})\|^2 \quad (25)$$

where $\hat{\mathbf{J}}_{\mathbf{g}} = \left[\hat{\mathbf{J}}_{\mathbf{g},x}, \hat{\mathbf{J}}_{\mathbf{g},y}, \hat{\mathbf{J}}_{\mathbf{g},z}, \hat{\mathbf{J}}_{\mathbf{g},sz}\right]^\top$. Unlike in Section 4.4, the calculation of each Jacobian is not identical due to the different normalisation procedure taken for each component. Given that we can split the partial derivative $\mathbf{J}_{\mathbf{g},x} = \mathbf{J}_x \odot \frac{\partial \mathbf{g}_{I,x}(\mathbf{p})}{\partial \mathbf{p}}$, we define the component specific

Jacobian, $\mathbf{J}_i \forall i \in \{x, y, z, sz\}$ as:

$$\mathbf{J}_x = \frac{\mathbf{g}_{I,y}(\mathbf{p})^2}{\left(\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2 + \mathbf{g}_{I,z}(\mathbf{p})^2\right)^{3/2}}$$

$$\mathbf{J}_y = \frac{\mathbf{g}_{I,x}(\mathbf{p})^2}{\left(\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2 + \mathbf{g}_{I,z}(\mathbf{p})^2\right)^{3/2}}$$

$$\mathbf{J}_z = \frac{\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2}{\left(\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2 + \mathbf{g}_{I,z}(\mathbf{p})^2\right)^{3/2}} \tag{26}$$

$$\mathbf{J}_{sz} = -\frac{\mathbf{g}_{I,z}(\mathbf{p})\left(\frac{\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2}{\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2 + \mathbf{g}_{I,z}(\mathbf{p})^2}\right)^{3/2}}{\mathbf{g}_{I,x}(\mathbf{p})^2 + \mathbf{g}_{I,y}(\mathbf{p})^2}$$

Now, given the definitions of the correct Jacobians per component, we can solve (25) as:

$$\Delta\mathbf{p} = \mathbf{H}^{-1}\hat{\mathbf{J}}_{\mathbf{g}}^{\top}\left[\hat{\mathbf{g}}_T(\mathbf{0}) - \hat{\mathbf{g}}_I(\mathbf{p})\right] \tag{27}$$

Given that the update in (27) is identical to that of (8), it would be trivial to formulate an inverse compositional form of this residual by following the steps described in Section 4.3.

5. Robust Nonrigid Alignment

Non-rigid registration is the term generally used to describe an alignment algorithm that utilises a non-rigid warp. A non-rigid warp is generally achieved via a motion model that allows for smaller scale local deformations than can be achieved under models such as affine or similarity. For example, Rueckert et al. [26] proposed free-form deformations (FFD) as a motion model that gives a smooth spline-based transition between neighbouring control points. However, due to the local nature of non-rigid alignment algorithms, they require many more parameters. In the case of FFDs there may be many thousands of parameters depending on the resolution of the FFD chosen. Unfortunately, due to the complexity of the parameter space, this causes Gauss- Newton algorithms such as those described in Section 4 to be infeasible. This is primarily due to the fact that the size of the Hessian matrix is defined by the number of parameters and a large Hessian matrix may be non-invertible under reasonable memory requirements.

Therefore, we augment the FFD algorithm given in [26] to use our similarity measure. The local transformation described by a FFD consists of a mesh of control points, $\phi_{i,j,k}$, separated by a uniform spacing. The FFD is then given as in [26], as a 3-D tensor product of 1-D cubic B-splines

$$\mathcal{W}(\mathbf{x}; \mathbf{p})_{local} = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\phi_{i+l,j+m,k+n} \tag{28}$$

where B_l describes the lth basis function of the B-spline, $B_0(u) = (1-u)^3/6$, $B_1(u) = (3u^3 - 6u^2 + 4)/6$, $B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6$, $B_3(u) = u^3/6$ and i, j and k are control point indices over the x, y and z axis respectively. We then perform a simple gradient descent algorithm, which terminates when a local minima is reached.

Typically, the total cost function for the FFD algorithm consists of a similarity term that depends in the data and a regularisation term that enforces smoothness on the local transformation. In this work we seek to improve the performance of the similarity term in the presence of systematic errors.

Let us assume that the parameters of the FFD are $\Phi = \{\phi_{i,j,k}\}$. We replace the normalised mutual information similarity measure $C(\Phi)_{\text{Similarity}}$ from the original FFD algorithm with our new robust similarity. For example, for the inner product similarity as described in Section 3.2.2 we define $C(\Phi)_{\text{IP}}$ as

$$C(\Phi)_{\text{IP}} = \tilde{\mathbf{g}}_I(\Phi)^{\top}\tilde{\mathbf{g}}_T(\mathbf{0}) \tag{29}$$

The parameters, Φ, can then be updated in gradient descent form as follows:

$$\Phi = \Phi + \mu\frac{\nabla_\Phi C(\Phi)_{\text{IP}}}{\|\nabla_\Phi C(\Phi)_{\text{IP}}\|} \tag{30}$$

where $\nabla_\Phi C(\Phi)_{\text{IP}}$ is the gradient of the similarity measure.

5.1. Numerical Stability

As discussed in [20], it is not possible to use normalised gradient fields directly due to discontinuities in differentiation. We thus regularise the normalised gradient fields using the technique presented in [23].

$$C(\Phi)_{\text{IP}} = \sum_{k\in\Omega}\frac{\mathbf{g}_I(\Phi)[k]^{\top}\mathbf{g}_T(\mathbf{0})[k] + \varrho\tau}{\|\mathbf{g}_I(\Phi)[k]\|_\varrho\|\mathbf{g}_T(\mathbf{0})[k]\|_\tau} \tag{31}$$

where $\|\cdot\|_* = \sqrt{\langle\cdot,\cdot\rangle + *^2}$ and Ω is the set of indices corresponding to the target image support. In this work ϱ and τ require only a single parameter, as opposed to the user specified regularisation values chosen in [23]. Explicitly, ϱ and τ are computed following an automatic choice based on total variation

$$\varrho = \frac{\eta}{V_I}\sum_{k\in\Omega_I}\mathbf{g}_I(\Phi)[k], \quad \tau = \frac{\eta}{V_T}\sum_{k\in\Omega_T}\mathbf{g}_T(\mathbf{0})[k] \tag{32}$$

where $\eta > 0$ is a parameter for noise filtering and V_* is the volume of interest in the image domain Ω_*.

5.2. Robustness Against Bias Fields

In this section we provide a formal proof of the robustness of our cost functions to bias field corruption. Consider an image signal M with no intensity inhomogeneities and a smooth signal Q, representing a multiplicative bias field [38, 39]. We assume Q to be constant within a small
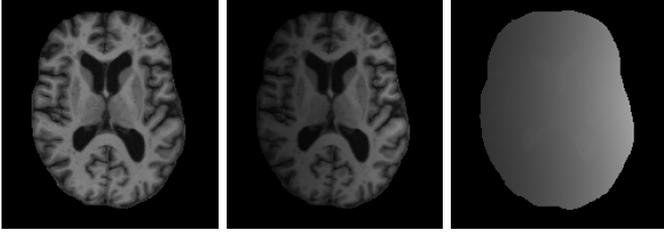
Figure 5: Axial view of a T1-weighted brain images utilised for intensity inhomogeneity simulation. Left: Original. Middle: With simulated bias field applied. Right: Bias field

neighbourhood, $N(k) = (\Delta k_x \Delta k_y \Delta k_z)$. Therefore, for Δk_x, we have

$$I(k) = M(k)Q(k) + \epsilon$$
$$I(k + \Delta k_x) = M(k + \Delta k_x)Q(k + \Delta k_x) + \epsilon \tag{33}$$

where ϵ is an additive error term. Given that Q is constant within N

$$
\begin{aligned}
\frac{\partial I(k)}{\partial x} &\approx \lim_{\Delta k_x \to 0} \frac{I(k + \Delta k_x) - I(k)}{\Delta k} \\
&\approx \lim_{\Delta k_x \to 0} \frac{Q(k)(M(k + \Delta k_x) - M(k))}{\Delta k} \\
&= Q(k)\frac{\partial M(k)}{\partial x}
\end{aligned}
\tag{34}
$$

Using (34), we show our proposed cost function to be robust to locally constant bias fields, since

$$
\begin{aligned}
&\frac{\frac{\partial I(k)}{\partial x}}{\sqrt{\left(\frac{\partial I(k)}{\partial x}\right)^2 + \left(\frac{\partial I(k)}{\partial y}\right)^2 + \left(\frac{\partial I(k)}{\partial z}\right)^2}} \\
&\approx \frac{Q(k)\frac{\partial M(k)}{\partial x}}{\sqrt{\left(Q(k)\frac{\partial M(k)}{\partial x}\right)^2 + \left(Q(k)\frac{\partial M(k)}{\partial y}\right)^2 + \left(Q(k)\frac{\partial M(k)}{\partial z}\right)^2}} \\
&\approx \frac{\frac{\partial M(k)}{\partial x}}{\sqrt{\left(\frac{\partial M(k)}{\partial x}\right)^2 + \left(\frac{\partial M(k)}{\partial y}\right)^2 + \left(\frac{\partial M(k)}{\partial z}\right)^2}}
\end{aligned}
\tag{35}
$$

Equations (33) — (35) are analogous for Δk_y and Δk_z. Therefore,

$$\frac{\nabla I(k)}{\|\nabla I(k)\|} \approx \frac{\nabla M(k)}{\|\nabla M(k)\|} \tag{36}$$

which demonstrates the invariance of our cost functions with respect to Q.

6. Experiments

We assessed the performance of our similarity measures within two separate experimental frameworks. For the LK experiments, we used data from the Visible Human project [40] and demonstrate the robustness of our proposed measures to both a simulated bias field and artificial occlusions.

For the non-rigid registration experiments, we demonstrate the robustness of our similarity measures to intensity inhomogeneities in the form of a bias field as well as simulated pathologies. The pathologies are introduced in the form of simulated brain tumours provided by the BraTS MICCAI 2012 challenge[1] image database. We also used MR brain images from the OASIS dataset [41] and traumatic brain injury MR images from Turku University Hospital.

6.1. Intensity Inhomogeneity Implementation

Robustness to intensity inhomogeneities relaxes the necessity of an explicit intensity correction step in the registration pipeline, which can be time consuming and a potential source of errors, especially for non-brain images. Any similarity measure that can accurately align images containing intensity inhomogeneities is of great benefit for data sources such as MRI data where bias field corruption is very common. Bias field corruption is a low-frequency and very smooth signal that corrupts MRI images, especially those produced by older MRI machines.

To introduce intensity inhomogeneities into the images, we simulate several two-dimensional complex-valued MRI sensitivity maps using a MATLAB tool[2]. For each image, we simulate the effect of 8 coils uniformly placed according to the software implementation. Then, we randomly select one of the 8 generated sensitivity maps as the final map S for the image. Since the sensitivity maps are two-dimensional, we apply them to every 2D slice of the image along the Z-axis in a weighted fashion. Hence, if we denote the original image as M, then the simulated image with intensity inhomogeneities I is constructed according to

$$I(\cdot, \cdot, z) = R\left(w(z) \otimes \|S(\cdot, \cdot, z)\| \otimes M(\cdot, \cdot, z)\right),$$

$$z \in [1, N_Z],$$

where N_z corresponds to the number of image slices in the Z direction, $R(\cdot)$ is the function that rounds the argument to the nearest integer, $\otimes$ is the voxelwise multiplication and $w(z)$ is given by

$$w(z) = 1 + \frac{10}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(z - \frac{N_z - 1}{2}\right)}{2\sigma^2}}.$$

For all the simulations we use $\sigma = 0.15 \cdot (N_z - 1)$. Figure 5 shows an example image with and without intensity inhomogeneities.

6.2. 3D Affine Registration Using LK

For the LK experiments, the data used was from the Visible Human project [40]. This data has an image structure that differs from other common 3D image sources such as MR images. Each voxel in an image is formed from

---

[1]http://www2.imm.dtu.dk/projects/BRATS2012/
[2]bigwww.epfl.ch/algorithms/mri-reconstruction
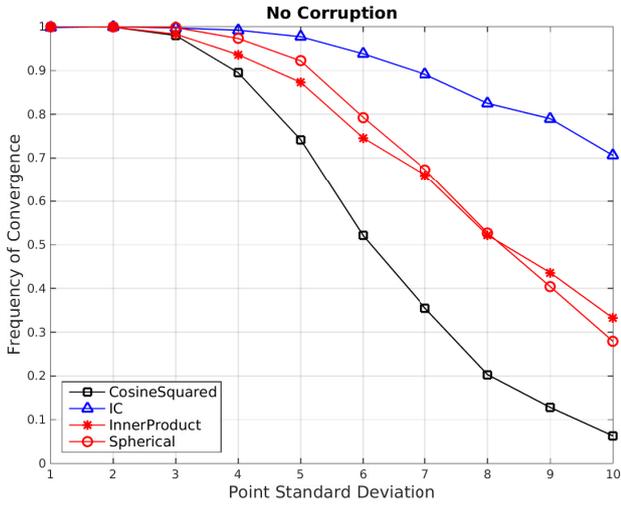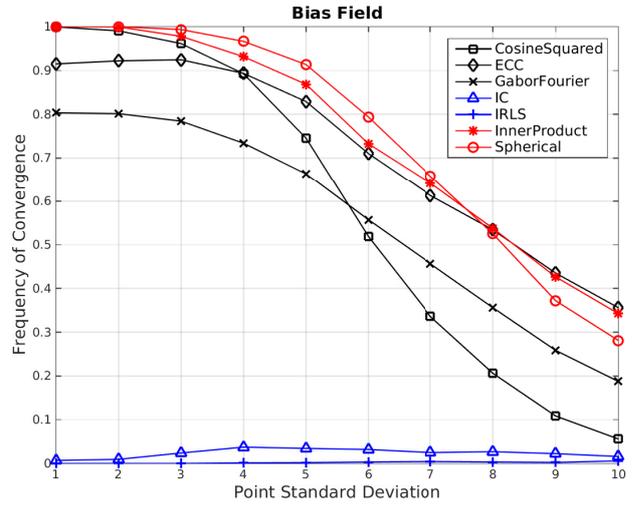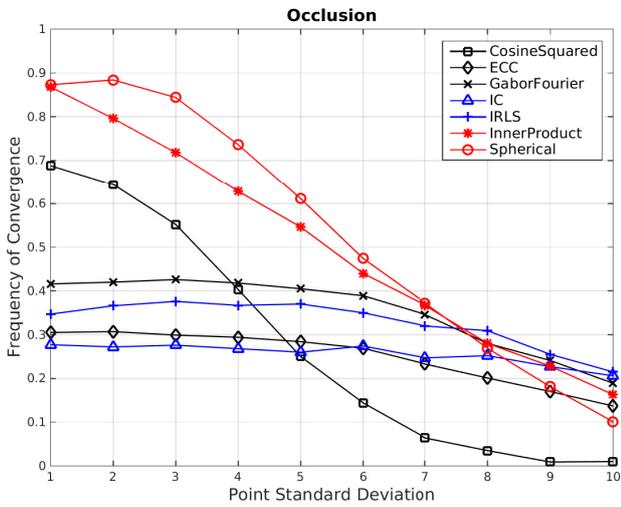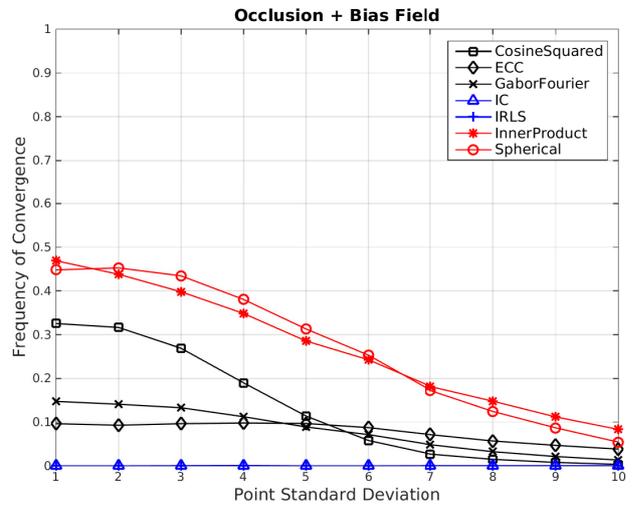
Figure 6: Average frequency of convergence vs Point standard deviation for the Visible Human data set. (a) Simulated Bias Field (b) Occlusions (c) Occlusions + Bias Field. CosineSquared: black-□. ECC: black-◇. GaborFourier: black-x. IC: blue-x. IRLS: blue-+. InnerProduct: red-*. Spherical: red-o
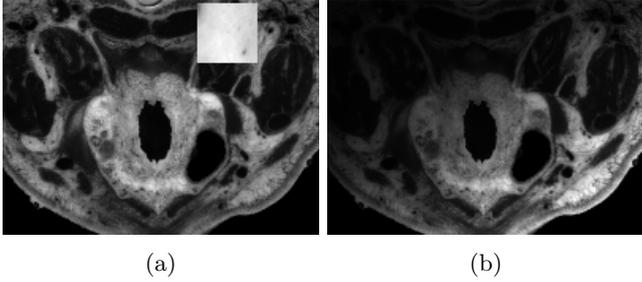
Figure 7: Examples of images from the Visible Human project [40] used in the LK experiments. (a) Artificially occluded image. The occlusion appears as the white square in the top right. (b) Image with simulated bias field.

physical slices that were taken from the body of a cadaver. Therefore, the gradient information across the x, y plane is incredibly rich as it represents a true 2D-image. However, the 3D nature of the data is still maintained as each image along the z-axis represents another slice acquired from the body.

In all experiments an affine motion model was used and performance was measured within an extension of the evaluation framework proposed in [5]. We used the Oral section from the Visible Human data set [40] as the target image. We selected 10 different regions of interest and parametrised the regions as a set of points representing the bounding cube of the region. These points were then perturbed using Gaussian noise of standard deviation σ. Using the affine warp defined between the original and perturbed points, we generate a distorted image. Then, given a warp estimate, we compute the new template points and calculate the root mean square error (RMSE) between the estimated and correct locations. The performance metric used to assess the algorithms is the average convergence rate for each fixed $\sigma = [1, 10]$, over each of the 10 regions of interest. An algorithm was considered to have converged if it had a final RMSE of less than 2.0 pixels after 30 iterations. For each template, 100 convergence tests were performed. Each image was smoothed using Gaussian smoothing with $\sigma = 2.0$ and kernel size $5 \times 5 \times 5$, before the calculation of derivatives. All algorithms were implemented using the inverse compositional form.

To provide a competitive assessment of our similarity measures, we extended recent state-of-the-art 2D algorithms for use with 3D images. We concentrated on algorithms that aim to provide robustness against outliers, particularly in the form of intensity inhomogeneities. Therefore, we provide comparisons against the enhanced correlation coefficient (ECC) [6] and the Fourier LK algorithm with Gabor filter banks (GaborFourier) [7]. As a baseline, we also compare against the standard LK algorithm and the iteratively re-weighted least squares algorithm (IRLS) also proposed by Baker et al. [2].

We also compare against the most related technique that utilises the cosine squared measure (CosineSquared) [20]. Our implementation of CosineSquared is equivalent to the Gauss-Newton methodology described within [20].

6.2.1. Experiments Without Corruption

In this subsection, we present our performance evaluation results obtained without applying any corruption to the 3D images. We compared the performance of the inverse-compositional LK algorithm (IC) with both forms of our algorithm, InnerProduct and Spherical and the related method CosineSquared. This experiment is designed as a baseline that presents the performance of robust measures in data that is known to contain no outliers.

As Figure 6a shows, the IC algorithm outperforms the other methods for this experiment. This result is unsurprising, as the distorted image is generated directly from the original image without any outliers. Since both of our proposed methods discard information in the form of the gradient magnitude, they inevitably perform worse than the LK algorithm. However, the difference between our two algorithms is negligible, which is expected given that they both discard the same amount of information. The larger deformations significantly decrease the performance of the CosineSquared algorithm. This is likely due to the bias created by squaring the inner product of the images.

6.2.2. Experiments With Corruption

In this subsection, we present three separate experiments: images with a simulated bias field, with an occlusion and with an occlusion and a simulated bias field. The bias field was generated as described in Section 6.1 and an example is shown in Figure 7b. Occluded sections were created synthetically by randomly placing image sections taken from another random area of the body, and putting them into every slice of the 3D image, as shown in Figure 7a.

Figure 6b shows that our proposed techniques are competitive with the state-of-the-art for bias field corruption. The LK and IRLS algorithms are not able to cope with the intensity variation caused by the bias field. GaborFourier copes reasonably well with this type of corruption due to the illumination invariant properties described in [7]. ECC performs very well, which is unsurprising as the enhanced correlation coefficient performs a normalisation of the image pixels, which reduces the effect of the bias field. The CosineSquared algorithm performs well for smaller deviations, but quickly diminishes in performance.

Figure 6c shows that our proposed similarity measures are also the most robust to occlusions. IRLS performs better under these situations as it is able to discard some of the outliers that bias the alignment. The normalisation step in ECC has no benefit in suppressing this sort of bias, and so it performs very similarly to the non-robust IC algorithm. GaborFourier still performs well as the Gabor filter banks suppress the contribution of the outliers. The CosineSquared performs well under smaller deformations but is heavily biased under large deformations as the squaring of the cosine fails to suppress the contribution of the occlusions.

12
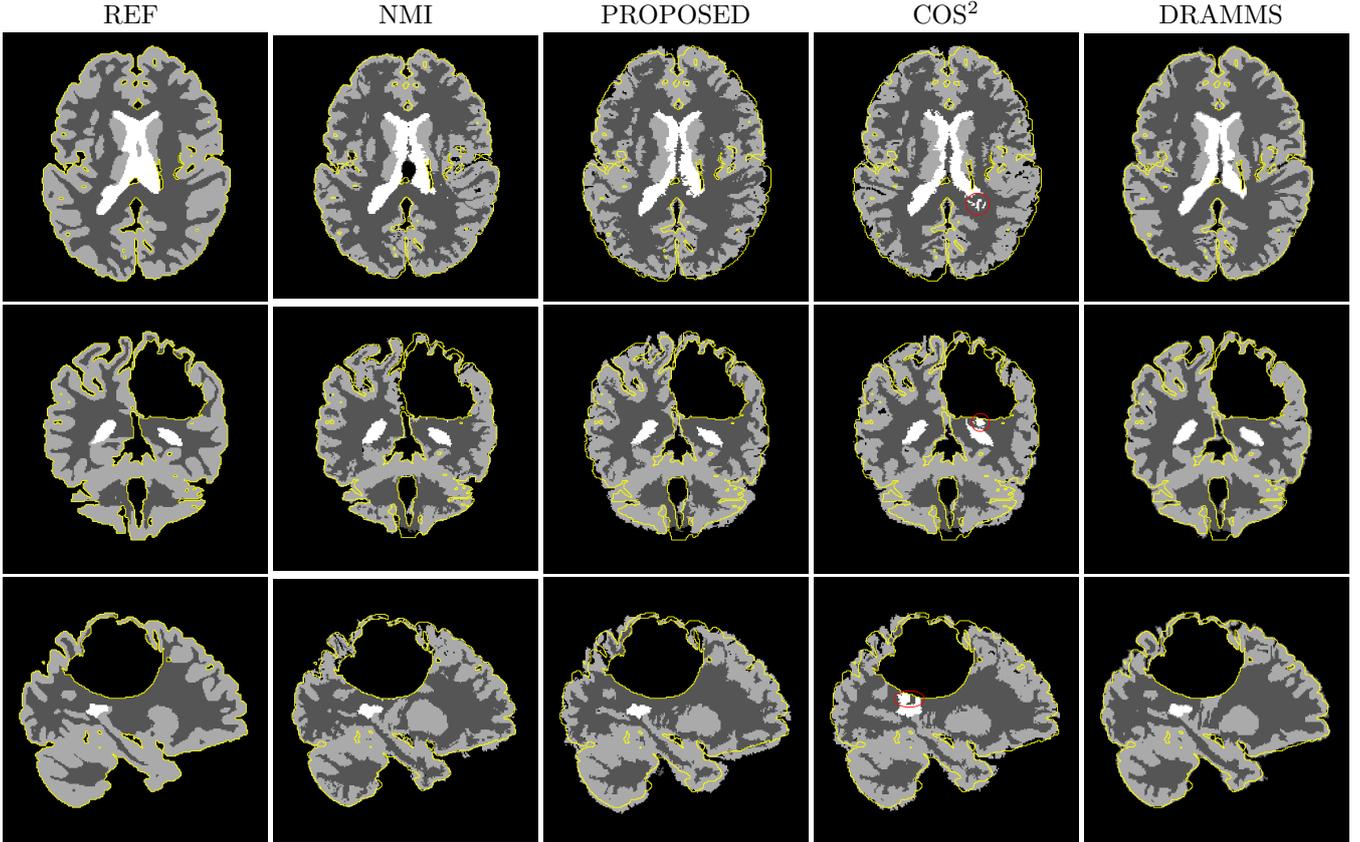
Figure 8: Reference and propagated labels using $\eta = 0.1$. REF denotes the reference labels, NMI is normalised mutual information, PROPOSED is the proposed similarity, COS^2 is Haber and Modersitzki [20], DRAMMS is the method of [27]. The first row is the axial view, second row is the coronal view and final row is the sagital view. Boundaries of the tumours and image are provided for visualisation.

Finally, in Figure 6d we see that even under occlusion and global illumination variation, our proposed measures perform with relatively high accuracy. This is a challenging experiment which demonstrates the power of our proposed similarity measure. Despite the large amount of outliers, our proposed measures are still able to perform accurate alignment with a higher success rate than any of the other algorithms considered.

6.3. 3D Non-Rigid Registration Using FFDs

As previously mentioned, we incorporate the proposed inner product measure into a B-Spline FFD algorithm [26]. For comparison, we also incorporate the cosine squared similarity [20] and normalised mutual information (NMI) [19] into our framework. We also compared our FFD-based non-rigid registration on images containing pathologies with a recent work, DRAMMS [27]. DRAMMS attempts to reduce matching ambiguities through a multi-scale and multi-orientation Gabor filters. These extracted filters are then weighted during registration using a function dubbed "mutual-saliency" designed to improve matching uniqueness. DRAMMS has been shown to be effective on a wide range of data sources. However we show that whilst not failing completely on images containing pathologies,

our FFD-based similarity measure yields superior performance.

We concentrate on two separate experiments, images with intensity inhomogeneities and images with pathologies. In all the conducted experiments within the FFD framework, we utilise the thin-plate bending energy of the deformation field with a weight of $\alpha = 0.002$ as a regularisation term and optimise using conjugate gradient descent. We use the same regularisation weight for every similarity since the empirical range of values that they take using our experimental image datasets are of very similar width for all of them.

6.3.1. MR Images with Intensity Inhomogeneities

Here we evaluate the performance of our proposed similarity measure against intensity inhomogeneities. This relaxes the necessity of an explicit intensity correction step in the registration pipeline (e.g. [42]), which can be time consuming and a potential source of errors, especially for non-brain images. To evaluate the proposed similarity measure for non-rigid registration, we perform the $30 \times 29 = 870$ pairwise registrations with control point spacings of 20, 10, 5 and 2.5mm, using the 35 original preprocessed T1-weighted MR brain images from the OASIS dataset [41]. We subsequently introduce different smooth intensity in-
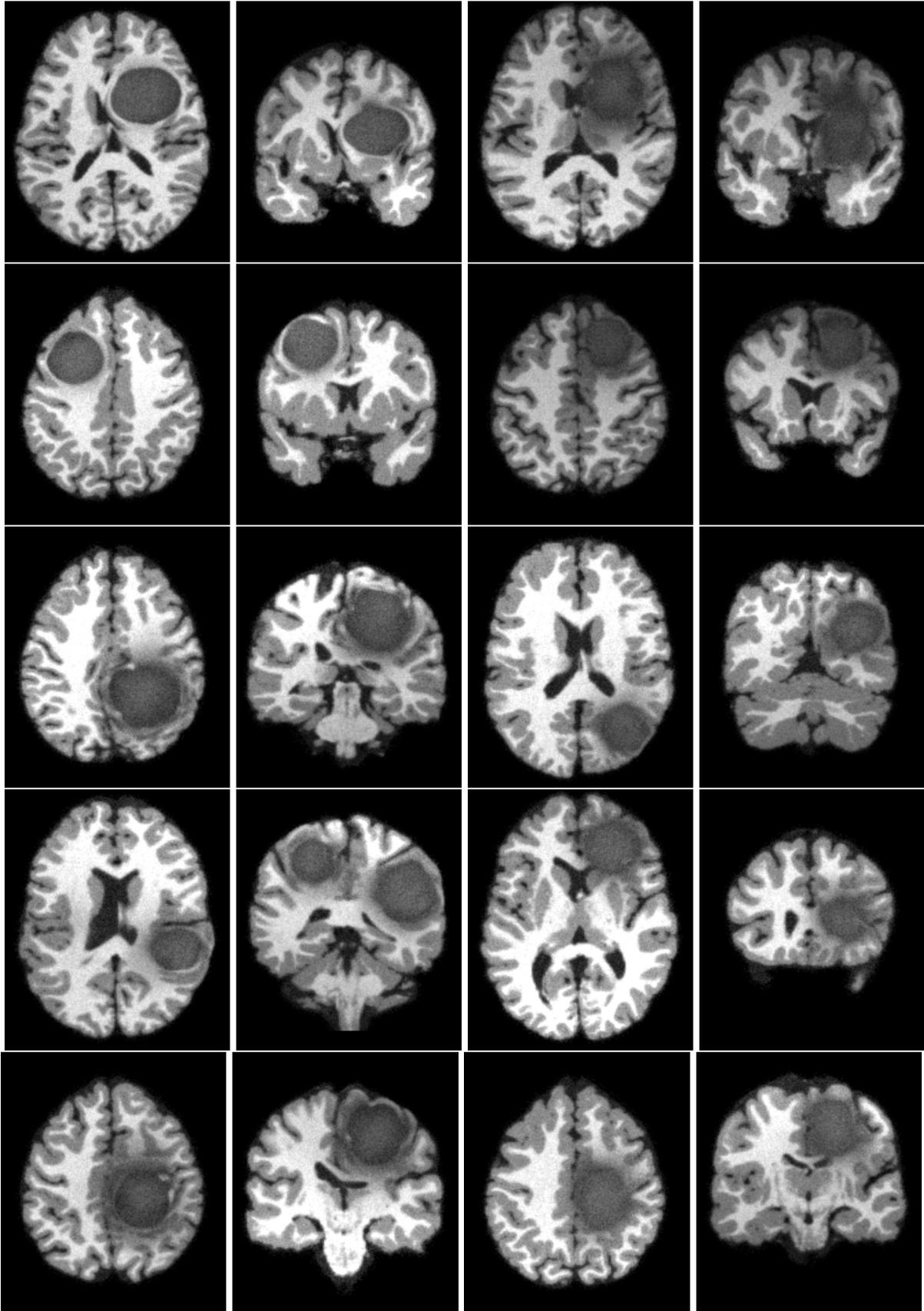
Figure 9: Examples images of a BraTS subject in axial and coronal view. First and second columns depict subjects with high-grade gliomas. Third and fourth columns depict subjects with low-grade gliomas.

NMI

NMI with bias field

Proposed similarity ($\eta = 0.1$)

Proposed similarity ($\eta = 0.1$) with bias field

Haber and Modersitzki [20]

Haber and Modersitzki [20] with bias field
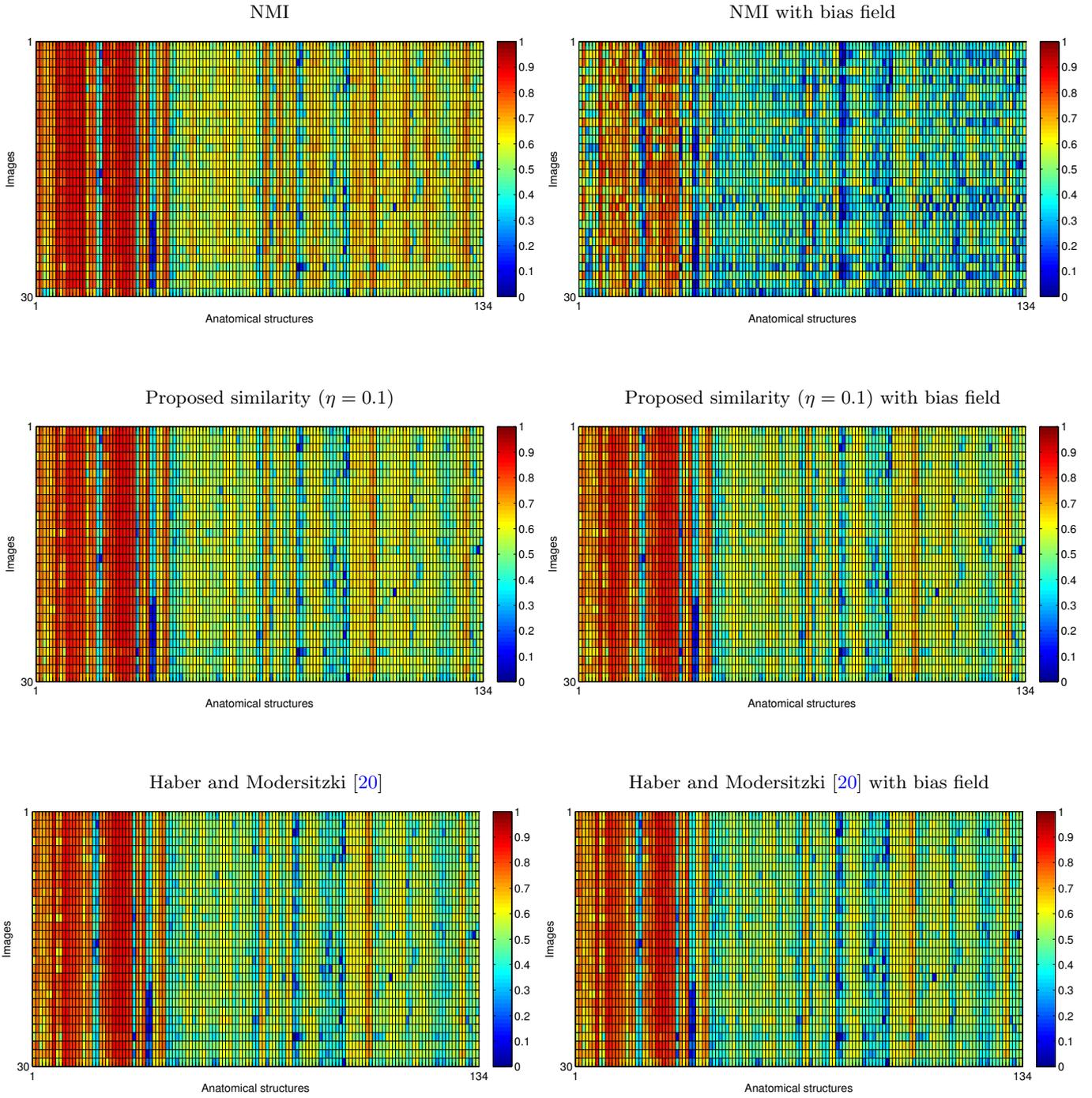
Figure 10: Mean pairwise similarity index per image per label for the OASIS data [41].

		WM	GM	CSF	Overall
NMI [19]		$78.1(78.1) \pm 0.03$	$79.0(78.0) \pm 0.03$	$88.4(88.9) \pm 0.02$	$81.8(81.3) \pm 0.02$
DRAMMS [27]		$73.7(74.2) \pm 0.02$	$79.0(79.5) \pm 0.03$	$78.4(81.2) \pm 0.11$	$77.1(78.6) \pm 0.05$
Proposed similarity	$(\eta = 0.1)$	$78.1(78.1) \pm 0.00$	$76.4(76.6) \pm 0.01$	$83.4(84.5) \pm 0.03$	$79.3(79.9) \pm 0.01$
Haber and Modersitzki [20]	$(\eta = 0.1)$	$74.2(74.4) \pm 0.01$	$75.1(75.0) \pm 0.01$	$80.9(82.2) \pm 0.03$	$76.8(77.2) \pm 0.02$
Proposed similarity	$(\eta = 1)$	$80.9(80.8) \pm 0.00$	$79.1(79.3) \pm 0.01$	$88.1(89.1) \pm 0.03$	$82.7(83.3) \pm 0.01$
Haber and Modersitzki [20]	$(\eta = 1)$	$81.2(81.2) \pm 0.00$	$79.6(79.7) \pm 0.01$	$88.4(89.0) \pm 0.03$	$83.1(83.7) \pm 0.01$

Table 1: **Images with pathology:** Mean (Median) overlap measures and standard deviation for white matter (WM), grey matter (GM) and ventricular cerebrospinal fluid (CSF) labels propagated using on-rigid registration.

(a) $\eta = 0.1$

(b) $\eta = 1$
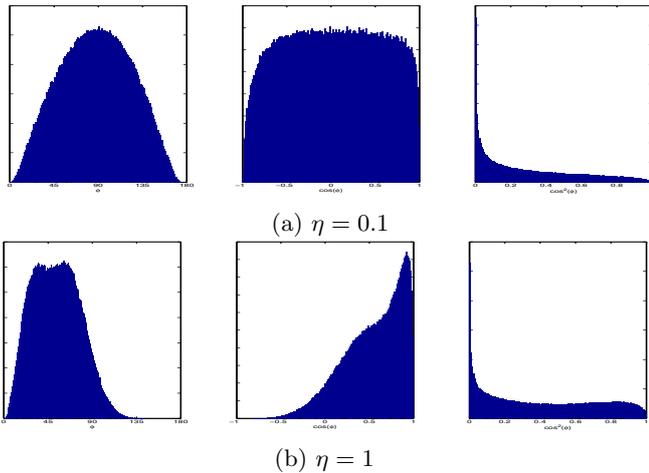
Figure 11: Angle ϕ and histograms of $\cos \phi$ and $\cos^2 \phi$ using (a) $\eta = 0.1$ and (b) $\eta = 1$ between a healthy subject and the BraTS simulated images in the tumour areas. The means are 0.031 and 0.296 respectively for $\eta = 0.1$ and 0.548 and 0.410 respectively for $\eta = 0.1$ $\eta = 1$.

homogeneities individually to all the images according to the procedure described in Section 6.1 and repeat the registrations again using the original images as target and the affected ones as source.

We compare the gradient-based similarity measures with noise parameter η set to 0.1 against NMI in their ability to produce a deformation field able to accurately propagate the manual segmentation labels. We measure the registration accuracy using the similarity index (SI), both for the original images and the images with bias field applied. We compute the mean and standard deviation of the SI values calculated on the propagated and reference labels for all 870 propagations. We differentiate between the 98 cortical and 36 subcortical labels. The results are shown in Figure 13. We observe that NMI performs well when there are no intensity inhomogeneities in the images. On the contrary, it is severely affected by the presence of intensity inhomogeneities. Conversely, both gradient-based similarity measures show similar performance for registrations with and without intensity inhomogeneities, demonstrating their robustness. Nevertheless, the proposed similarity measure performs slightly better that cosine squared, and

the differences are statistically significant according to a paired Wilcoxon signed rank test ($p < 10^{-5}$). To complement the analysis, we show the same results for each of the images and each label in Figure 10. It is important to note that in the case where no intensity inhomogeneities are present, the proposed method has a lower performance than NMI. The conducted analyses suggest that, when using normalised gradient fields, the registration of MR images is more difficult than the alignment of scans from other imaging modalities as in [21, 22, 23]. We observe that in the particular case of MR brain images, the discrimination between noise and structure related gradients is very challenging, especially in cortical areas.

6.3.2. MR Images with Pathologies

Registration of images depicting pathology is a challenging procedure, since the images may exhibit strong structural differences that cannot be matched. Here, we show that our similarity measure is capable of handling images with areas of mismatches, e.g., areas of pathology, without any prior knowledge nor any subsequent correction step.

We use a dataset of 10 simulated MR images of the brain depicting tumours. These images are taken from the BraTS MICCAI 2012 challenge[3]. Half of these images show high grade gliomas and the other half has low grade ones. The images are labelled into white matter (WM), gray matter (GM), cerebrospinal fluid (CSF) and 2 further labels for the tumour areas. All the images are skull stripped, and have $256 \times 181 \times 256$ voxels with an isotropic resolution of 1mm. A visualisation of all the subjects from this dataset is provided in Figure 9. For a quantitative evaluation, a labelled image of a healthy subject is registered to all 10 images in our BraTS dataset using NMI and both normalised gradient field based similarity measures. The registrations using either the cosine or cosine squared similarity measure were run in two settings. In the first setting, the noise parameter η is set to 0.1. In the second, we set the value of η to 1. We measure registration accuracy using SI over three labels, namely WM, GM and CSF. We ignore the two available tumour labels as there is

---

[3]http://www2.imm.dtu.dk/projects/BRATS2012/

16

no equivalent in the healthy scan. A good overlap for non-tumour labels is an indicator that the similarity measure is not biased by the presence of pathology. We also include the registration result of DRAMMS [27], which shows that FFD registration is still highly competitive with more recent state-of-the-art works. Overlap results for non-rigid registration are shown in Table 1. It can be observed that for the case where $\eta = 0.1$, the proposed similarity measure yields a worse alignment than NMI ($p = 0.0098$). However, when compared to the cosine squared similarity, a much better result is obtained for the proposed similarity measure ($p = 0.002$). Further visual results for this comparison are shown in Figure 8. The main areas where the registration using cosine squared similarity measure is affected by the tumour presence are highlighted by a red ellipse. Although Table 1 suggests that higher values of the parameter η provide superior overlap measures this is not necessarily the case. Given Equation (31) and Equation (32), we can clearly see that η has the effect of suppressing the contribution of the gradients within the similarity measure. Therefore, η can be thought of as a filtering parameter that will benefit any normalised gradient field similarity measures by decreasing the effect of noise. However, as demonstrated in Figure 11, larger values of η also effect the distribution of ϕ in the occluded areas. As η is increased, the gradient values are suppressed towards 0 and thus the distribution of $\cos \phi$ ceases to approximate a uniform distribution. Therefore, although $\eta = 1$ provides superior results for the experiment in Table 1, we focus our results on $\eta = 0.1$ whereby our assumption of the uniform distribution of outliers is not violated. This is clearly demonstrated by Figure 11 for the BraTS simulated images in the tumour areas when $\eta = 0.1$. Figure 11 differs from Figure 2 in its application of the noise filtering as no filtering is applied in Figure 2. This can be seen as equivalent to $\eta = 0$ which was used to validate that the difference between two visually dissimilar areas does indeed approximate a uniform distribution for the cosine.

In addition to the BraTS experiments, we perform registrations using NMI, the proposed similarity measure and cosine squared similarity measure on traumatic brain injury (TBI) data. The imaging data was acquired at Turku University Hospital, Finland in the course of the TBIcare project[4]. One image corresponds to a baseline scan and the other is the follow-up scan taken after 4 months and 19 days at Turku University Hospital. Both images have $176 \times 240 \times 256$ voxels with an isotropic resolution of 1mm.

We utilise the baseline image as target and the follow-up image as source. For the gradient-based similarity measures we set η to 0.1. Visual results are given in Figure 12. The main observation is that, in contrast to NMI and the cosine squared similarity measure, the proposed measure is able to recover most of the underlying changes in shape within the pathology area (as pointed out with a red arrow). This is because the registration inside the area of
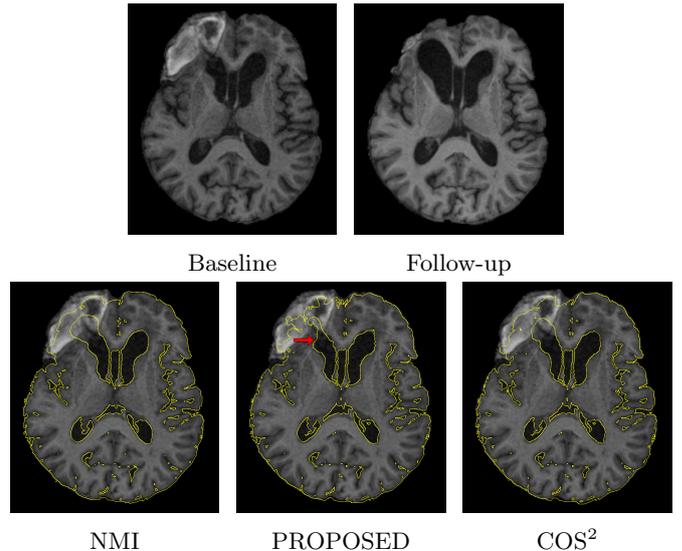
Figure 12: Visual comparison of the proposed similarity (PROPOSED), normalised mutual information (NMI) [18] and cosine squared (COS^2) [20] for TBI data. Top row: Baseline and follow-up images. Second row: Transformed follow-up isolines of the registration results overlaid on the baseline image.
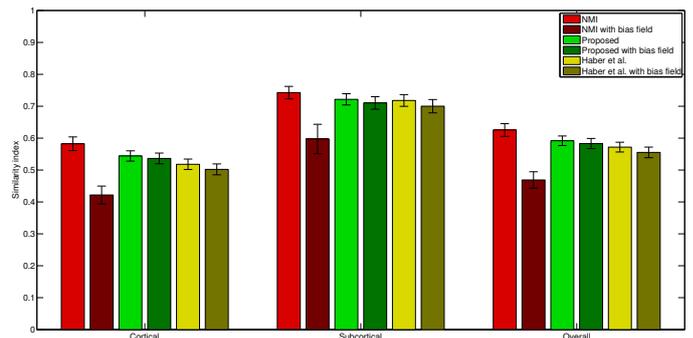
Figure 13: Mean similarity index and standard deviation over cortical and subcortical labels for all $30 \times 29 = 870$ registrations in the OASIS dataset [41].

pathology affect the value of both NMI and cosine squared similarity measure significantly, causing the optimisation to converge very quickly. On the other hand, the proposed similarity is almost unaffected by the forces in the area of pathology, allowing the optimisation to continue until good alignment is achieved.

7. Conclusion

In this work, we have provided a set of novel global similarity measures based on the cosine of the orientation between normalised gradients. We have shown that these measures are theoretically robust to a very general class of outliers that includes intensity inhomogeneities and systematic errors such as occlusions. We have empirically verified that our measures are robust for various sources of brain MRI data.

---

[4] http://www.tbicare.eu

We have provided a proof-of-concept investigation as to the effectiveness of our similarity measures within the popular Lucas-Kanade algorithm. Despite the popularity of Lucas-Kanade algorithms for 2D images, little work has been done on its performance for voxel data. In fact, to the best of our knowledge, no previous work exists that extends the current state-of-the-art Lucas-Kanade algorithms in to 3D. We provide a thorough experimental analysis on both our proposed measures and extensions of state-of-the-art 2D methods in Section 6.

We also embedded our similarity measures within a popular non-rigid alignment framework based on free-form deformations. We demonstrated the effectiveness and robustness of our similarity measure on images with simulated bias fields and on simulated pathological images. We showed superior robustness in these scenarios compared to NMI and the cosine squared measure of Haber and Modersitzki [20]. We also demonstrate superior performance on images containing pathologies when compared against DRAMMS [27].

The main contribution of this paper is that our similarity measures relax the need for bias field correction, which can be time consuming and prone to errors. They can also register images in the presence of pathologies, since they do not rely on any particular deformation model and do not require segmentation of outliers. Our similarity measures are very efficient and simple to compute and compare very favourably with current state-of-the-art methodologies.

Acknowledgments

References

[1] B. D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of the 7th international joint conference on Artificial intelligence, 1981.

[2] S. Baker, R. Gross, T. Ishikawa, I. Matthews, Lucas-Kanade 20 Years On: A Unifying Framework: Part 2, Tech. Rep., 2003.

[3] N. Dowson, R. Bowden, Mutual Information for Lucas-Kanade Tracking (MILK): An Inverse Compositional Formulation, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (1) (2008) 180–185.

[4] S. Lucey, R. Navarathna, A. B. Ashraf, S. Sridharan, Fourier Lucas-Kanade Algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (6) (2013) 1383–1396.

[5] S. Baker, I. Matthews, Lucas-Kanade 20 Years On: A Unifying Framework, Internation Journal on Computer Vision 56 (3) (2004) 221–255.

[6] G. D. Evangelidis, E. Z. Psarakis, Parametric image alignment using enhanced correlation coefficient maximization, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (10) (2008) 1858–1865.

[7] A. B. Ashraf, S. Lucey, T. Chen, Fast image alignment in the Fourier domain, in: CVPR, 2480–2487, 2010.

[8] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, M. Unser, Realistic analytical phantoms for parallel Magnetic Resonance Imaging, IEEE Transactions on Medical Imaging 31 (3) (2012) 626–636.

[9] F. J. P. Richard, A new approach for the registration of images with inconsistent differences, in: ICPR, 649–652, 2004.

[10] E. I. Zacharaki, D. Shen, S. Lee, C. Davatzikos, ORBIT: a multiresolution framework for deformable registration of brain tumor images., IEEE Transactions on Medical Imaging 27 (8) (2008) 1003–1017.

[11] M. Hachama, A. Desolneux, F. J. P. Richard, Bayesian Technique for Image Classifying Registration, IEEE Transactions on Image Processing 21 (9) (????) 4080–4091.

[12] X. Liu, M. Niethammer, R. Kwitt, M. McCormick, S. Aylward, Low-Rank to the Rescue – Atlas-Based Analyses in the Presence of Pathologies, in: Image Analysis and Processing, Springer International Publishing, Cham, 97–104, 2014.

[13] C. Studholme, C. Drapaca, B. Iordanova, V. Cardenas, Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change., IEEE Transactions on Medical Imaging 25 (5) (2006) 626–639.

[14] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, P. Suetens, Nonrigid Image Registration Using Conditional Mutual Information, IEEE Transactions on Medical Imaging 29 (1) (2010) 19–29.

[15] X. Zhuang, S. Arridge, D. J. Hawkes, S. Ourselin, A nonrigid registration framework using spatially encoded mutual information and free-form deformations., IEEE Transactions on Medical Imaging (c) (2011) 1–10.

[16] G. Tzimiropoulos, S. Zafeiriou, M. Pantic, Robust and efficient parametric face alignment, in: ICCV, IEEE, 1847–1854, 2011.

[17] S. Baker, R. Patil, K. M. Cheung, I. Matthews, Lucas-kanade 20 years on: Part 5, Robotics Institute, Carnegie Mellon University, Tech.Rep.CMU-RI-TR-04-64 .

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

[19] C. Studholme, D. L. G. Hill, D. J. Hawkes, An overlap invariant entropy measure of 3D medical image alignment., Pattern Recognition 32 (1) (1999) 71–86.

[20] E. Haber, J. Modersitzki, Intensity Gradient Based Registration and Fusion of Multi-modal Images, in: MICCAI, Springer Berlin Heidelberg, Berlin, Heidelberg, 726–733, 2006.

[21] S. Heldmann, S. Zidowitz, Elastic registration of multiphase CT images of liver., in: D. R. Haynor, S. e. b. Ourselin (Eds.), Proc. SPIE, 2009.

[22] T. Lange, N. Papenberg, S. Heldmann, J. Modersitzki, B. Fischer, H. Lamecker, P. M. Schlag, 3D ultrasound-CT registration of the liver using combined landmark-intensity information., IJCARS 4 (1) (2009) 79–88.

[23] J. Ruhaak, L. Konig, M. Hallmann, N. Papenberg, S. Heldmann, H. Schumacher, B. Fischer, A fully parallel algorithm for multimodal image registration using normalized gradient fields, ISBI (2013) 572–575.

[24] S. Pszczolkowski, Z. Stefanos, C. Ledig, D. Rueckert, A Robust Similarity Measure for NonRigid Image Registration With Outliers, in: ISBI, IEEE, 2014.

[25] S. Pszczolkowski Parraguez, Fast and robust methods for nonrigid registration of medical images, Ph.D. thesis, Imperial College London, 2014.

[26] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, D. J. Hawkes, Nonrigid registration using free-form deformations: application to breast MR images, IEEE Transactions on Medical Imaging 18 (8) (1999) 712–721.

[27] Y. Ou, A. Sotiras, N. Paragios, C. Davatzikos, DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting, Medical image analysis 15 (4) (2011) 622–639.

[28] G. D. Hager, P. N. Belhumeur, Efficient region tracking with parametric models of geometry and illumination, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (10) (1998) 1025–1039.

[29] M. J. Black, A. D. Jepson, EigenTracking: Robust matching and tracking of articulated objects using a view-based representation, in: ECCV, Springer-Verlag, Berlin/Heidelberg, 329–342, 1996.

[30] J. P. W. Pluim, J. B. A. Maintz, M. A. Viergever, Mutual-information-based registration of medical images: a survey, IEEE Transactions on Medical Imaging 22 (8) (2003) 986–1004.

[31] T. Butz, J.-P. Thiran, Affine Registration with Feature Space Mutual Information, in: MICCAI, Springer Berlin Heidelberg, Berlin, Heidelberg, 549–556, 2001.

[32] H. Livyatan, Z. Yaniv, L. Joskowicz, Gradient-based 2-D/3-D rigid registration of fluoroscopic X-ray to CT, IEEE Transactions on Medical Imaging 22 (11) (2003) 1395–1406.

[33] M. Droske, M. Rumpf, A Variational Approach to Nonrigid Morphological Image Registration, SIAM Journal on Applied Mathematics 64 (2) (2004) 668–687.

[34] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou, T. Stathaki, Robust FFT-Based Scale-Invariant Image Registration with Image Gradients, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (10) (2010) 1899–1906.

[35] S. Baker, I. Matthews, Equivalence and efficiency of image alignment algorithms, in: CVPR, IEEE, 1090–1097, 2001.

[36] I. Matthews, S. Baker, Active appearance models revisited, International Journal of Computer Vision .

[37] G. Papandreou, P. Maragos, Adaptive and constrained algorithms for inverse compositional Active Appearance Model fitting, in: CVPR, ISSN 1063-6919, 1–8, 2008.

[38] J. G. Sled, A. P. Zijdenbos, A. C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, IEEE Transactions on Medical Imaging 17 (1) (1998) 87–97.

[39] C. R. Meyer, P. H. Bland, J. Pipe, Retrospective correction of intensity inhomogeneities in MRI, IEEE Transactions on Medical Imaging 14 (1) (1995) 36–41.

[40] V. Spitzer, M. J. Ackerman, A. L. Scherzinger, D. Whitlock, The visible Human Male: A Technical Report, Journal of the American Medical Informatics Association 3 (2) (1996) 118–130.

[41] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, R. L. Buckner, Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults, Journal of cognitive neuroscience 19 (9) (2007) 1498–1507.

[42] N. Tustison, B. Avants, P. Cook, Y. Zheng, A. Egan, P. Yushkevich, J. Gee, N4ITK: improved N3 bias correction, IEEE Transactions on Medical Imaging 29 (6) (2010) 1310–20.