Enhancing the Multivariate Signal of $[^{15}O]$water PET Studies With a New Non-Linear Neuroanatomical Registration Algorithm

Ulrik Kjems, Stephen C. Strother, Jon Anderson,
Ian Law, Lars Kai Hansen

Abstract— **This paper addresses the problem of neuro-anatomical registration across individuals for functional $[^{15}O]$water PET activation studies. A new algorithm for 3D non-linear structural registration (warping) of MR scans is presented. The method performs a hierarchically scaled search for a displacement field maximizing one of several voxel similarity measures derived from the two dimensional histogram of matched image intensities, subject to a regularizer that ensures smoothness of the displacement field. The effect of the non-linear structural registration is studied when it is computed on anatomical MR scans and applied to co-registered $[^{15}O]$water PET scans from the same subjects; in this experiment a study of visually guided saccadic eye movements.**

The performance of the non-linear warp is evaluated using multivariate functional signal and noise measures. These measures prove to be useful for comparing different inter-subject registration approaches, e.g. affine versus non-linear.

A comparison of 12-parameter affine registration versus non-linear registration demonstrates that the proposed non-linear method increases the number of voxels retained in the cross-subject mask. We demonstrate that improved structural registration may result in an improved multivariate functional signal-to-noise ratio. Furthermore registration of PET scans using the 12-parameter affine transformations that align the co-registered MR images does not improve registration compared to 12-parameter affine alignment of the PET images directly.

Keywords— **Non-linear warping, stereo-tactic registration, inter-subject registration, voxel similarity measures.**

I. Introduction

When investigating functional activation patterns in multiple subjects, a basic problem is that of comparing and averaging the functional activation images across subjects. Apart from rigid-body translation and rotation, the inter-subject registration problem is essentially how to find a three dimensional deformation of a subject's anatomy to match a template anatomy. The deformation ideally represents a *correspondence map*, between the observed image and the template image, to the extent that such a map exists.

Multiple scans from one or more sensor modalities of the same subject may be automatically registered quite accurately using rigid-body rotations and translations, involving 6 parameters in the transformation procedure. In [1] it is shown how intra-subject registration of MRI-MRI and PET-MRI scans can be achieved with good accuracy across the entire volume ($\sim$1–2 mm) using the Automated Image Registration (AIR) method as described in [2], [3], [4], [5].

In contrast to this, inter-subject registration is more complex in terms of the number of parameters necessary: Ultimately, the correspondence map consists of three parameters for each voxel, representing the x, y and z location of the corresponding voxel in a template image typically defined in the standard Talairach stereotactic space [6]. Automated inter-subject registration procedures typically overcome this parameter estimation problem by parameterizing the deformation field in a lower dimension. An example of this is the 12-parameter transformation supported by the AIR package [2], [3], [4] which recently has been extended to include 5th order polynomial transformations with up to 168 parameters [5].

This paper compares the resulting functional activation patterns achieved when applying non-linear registration techniques across subjects to those achieved using conventional 12-parameter affine[1]

U. Kjems and L. K. Hansen are with the Department of Mathematical Modelling, Technical University of Denmark. E-mail: uk@imm.dtu.dk

S. C. Strother and J. Anderson are with the PET Imaging Service, VA medical center, Minneapolis. S. C. Strother is also with the Radiology, Neurology and Health Informatics Depts., University of Minnesota, Minneapolis

I. Law is with the Department of Neurology, National University Hospital of Denmark, Copenhagen

[1]An affine transformation is one that preserves straight parallel lines. It is conveniently expressed as a matrix multiplication of a 4 by 4 transformation matrix with a homogeneous coordinate vector [7].

transformations. The inter-subject functional variation (e.g. the variance across subjects in $[^{15}\mathrm{O}]$water PET studies) is composed of a mixture of measurement noise, anatomical variation and functional variation, e.g. variation in how a given subject's activation pattern is linked to the anatomy. Since the inter-subject structural registration is limited to a certain spatial scale, we need to choose the degree to which the structural registration should be performed. The image resolution together with the functional variation between subjects set a lower bound on the scale at which the anatomical registration will be useful. It may be that the effort of performing a detailed non-linear structural registration is not rewarded by an increase in the extracted functional activation signal, because the non-linear registration operates at a smaller scale than the functional differences that exist within a set of normal subjects. Furthermore, the optimal functional registration is not necessarily the most accurate anatomical registration. With this in mind, we wish to know whether our proposed non-linear registration algorithm leads to an "improvement" in the measured brain activity map, as opposed to using a simpler conventional method, e.g. 12-parameter affine transformations.

In the non-linear registration procedure presented different voxel similarity measures may be applied and it is not evident beforehand which of these measures is optimal. Furthermore, the algorithm uses a set of regularizing parameters, which control the smoothness of the displacement field that is obtained, and we need to consider how to find an optimal setting for such parameters.

In order to address the above problems, we have evaluated the anatomical registration by examining how well the corresponding functional data have been registered and how the functional activation signal is improved. We assume that the optimal registration of the functional images is the one that maximizes the multivariate separation of the functional activation signal (i.e. separation of the means of the baseline and activated state conditions) for a group of subjects. We do not assume that this registration is identical to the optimal registration from an anatomical point of view. In fact, the results of this experiment indicate that the functional $[^{15}\mathrm{O}]$water PET signal is *degraded* if we force the non-linear registration to work at the highest level of detail. This is due to two problems: firstly, there is no perfect relation between brain anatomy and brain function, and secondly there exists no perfect relation between individual subjects' anatomies.

Using the approach of letting the strength of the functional signal be an indicator of the quality of the anatomical registration, we are able to directly compare different registration techniques, voxel similarity measures and regularization parameter settings.

The rest of this paper is arranged as follows. The non-linear registration algorithm is described in Section II. In Section III we describe the functional data set used for the alignment experiments, and in Section IV the analysis methods used for examining the functional signal space are described and the results of the analysis are discussed.

The method described here is based on a 3D extension of the approach described in [8] although several extensions have been added. The method is also similar in construction to the ANIMAL algorithm developed by Collins *et al.* [9], [10] and the methods described in [11], [12]. The abundant normalization methods can be categorized into "intensity based", which compute a deformation by volumetric intensity similarity measures, and "model based" where a geometrical surface model is adapted to the volume data. The method presented here is intensity based. A considerable amount of work exists based on both classes of non-linear deformations of human brain anatomy—most of the work to date has investigated the anatomical variation between individuals [13], [14], [15], [16], [17], [18], [19]. Few authors have investigated the extent to which functional inter-subject variation in $[^{15}\mathrm{O}]$water PET scans can be reduced using anatomical information [20].

The major difference between this method and previously published methods lies in the regularization applied and in the evaluation of the voxel similarity measures. Instead of optimizing the similarity measure locally [13], this method performs the optimization based on the change in the global similarity measure subject to regularizing constraints. Furthermore, the method is optimized to yield fast convergence, so that a full non-linear registration of a high-resolution $256 \times 256 \times 160$ volume with a $64 \times 64 \times 40$ vector field can be computed in 25 minutes on a 200MHz R4400 processor. The techniques for evaluation of global change of similarity measures presented below may also be adapted to warp-algorithms that apply different strategies for searching for the optimal displacement field, e.g. the wavelet based method of [21].

II. The Non-linear Registration Algorithm

A volume of displacement vectors Φ overlying the template volume is computed. These vectors are computed on a grid which is coarser than the MR image (in this experiment 4 times coarser, i.e. one vector per $4 \times 4 \times 4$ block of MRI voxels) and represent local displacements that make the subject image volume similar to the template image. The vector field is computed in a hierarchical manner: The iterations are performed using three resolution steps (1/4, 1/2 and full resolution) of both subject and template MRI images. The displacement field increases it's resolution in a likewise manner, at all times using one displacement vector for each 4-cube of voxels. The resolution is increased whenever the similarity measure has converged, interpolating the displacement vectors using tri-linear interpolation before recommencing the iterations. Once the highest resolution (i.e. the original MR resolution) is reached and the displacement field has converged, a resampled subject MRI volume (matching the template) may be found by evaluating the displacement of each template voxel using tri-linear interpolation between the 8 local surrounding vectors on the grid of displacement vectors.

The details of the algorithm are explained using a pseudo-C description in Table I. We have had best results when the vector field is four times coarser than the template image, i.e., each vector is associated with the $4 \times 4 \times 4$ cube of nearest voxels. Three resolution steps have been sufficient for all scans tested when the scans have been prealigned using a 12-parameter affine transformation prior to the non-linear warp. We have used our own implementation[2] of the AIR method [1], [3], [4] for this step.

Let $\mathcal{A}(\boldsymbol{v})$ and $\mathcal{B}(\boldsymbol{v})$ be the template and subject image volumes, respectively, indexed by the vector $\boldsymbol{v} = (p,q,r) \in \mathcal{G}$, consisting of discretely sampled and (8 bit integer) quantized voxel intensities a and b. The deformation field at iteration n, denoted Φ^n, consists of vectors Φ^n_{ijk} with the coarser grid indexed by (i,j,k). The template voxel $\mathcal{A}(\boldsymbol{v})$ is subject to a shift in coordinates so that $\mathcal{A}(\boldsymbol{v})$ is mapped to $\mathcal{B}(\boldsymbol{v}+\Phi^n_{\boldsymbol{v}})$, where $\Phi^n_{\boldsymbol{v}}$ is obtained by trilinear interpolation of the 8 surrounding displacement vectors on the coarser sampled grid of vectors Φ^n_{ijk}.

Different assumptions about the exact nature of this mapping may be made depending on the sources of image volumes $\mathcal{A}$ and $\mathcal{B}$. If the scans are recorded using identical acquisition sequences on the same scanner, it may be fair to assume a linear mapping of the voxel intensities e.g. the correlation voxel similarity measure. Such an approach may not be viable if the tissue contrasts differ. In such cases a more general mapping should be assumed—e.g. the mutual information (MI) similarity measure. MI has gathered considerable interest for inter-modality image registration, see e.g. [22], [23], [24], [25], [26]. Variations

[2]This implementation is functionally equivalent to the AIR package.

TABLE I

THE OUTLINE OF THE WARPING ALGORITHM.

```
for resolution = quarter, half, full do

       subsample the template and subject volumes
       to current resolution using Hann weighted
       sinc sub-sampling kernel.

       Interpolate displacement field so that
       vectors exist for each 4 x 4 x 4 block of
       voxels in the subsampled template image.

       repeat

              Compute cost function look-up table
              T(a,b) from current histogram h^n(a,b).

              Update elements Φ_ijk for which
              i+j+k is even maximizing the
              estimated decrease in the global cost
              function
```

$$\Phi^n_{ijk} = \underset{\phi \in S_{ijk}}{\text{minarg}} \; \Delta\hat{C}_\phi$$

```
              Update elements Φ_ijk for which
              i+j+k is odd.

              Apply low-pass filter to Φ^n.

              Compute new histogram of matched image
              intensities h(a,b), and compute the
              global similarity measures from this.

       until (selected similarity measure has
       converged^a )

end for
```

[a] Convergence is assumed either when the similarity measure no longer changes (D is evaluated directly from $h(a,b)$) or when a maximum number of iterations have been performed (8 iterations were enough in all cases tested).

in tissue contrast *between* the volumes should not be confused with variation *within* a single volume typically caused by field inhomogeneities—such effects are presumed to have been corrected for prior to the registration process.

A. Iteration Procedure

At each iteration step, vector estimates are updated by searching for the largest possible estimated decrease in a global cost function, which is assumed to take the form

$$C(\Phi^n) = D\{\mathcal{A}(\boldsymbol{v}), \mathcal{B}(\boldsymbol{v}+\Phi^n_{\boldsymbol{v}})\}_{\boldsymbol{v}\in\mathcal{G}} + R(\Phi^n), \quad (1)$$

with $D(a,b)$ being a function measuring the similarity of a set of matched voxel intensities a and b, and $R(\Phi^n)$ is a regularizing constraint representing our prior knowledge of the nature of the displacement field, i.e. that it is smooth. We have applied a simple Gaussian Markov random field [27] prior consisting of the sum of squared differences of the 6 nearest neighboring vectors indexed by $i'j'k' \in \mathcal{N}_{ijk}$

$$R(\Phi^n) = \frac{\alpha}{2} \sum_{ijk} \sum_{i'j'k' \in \mathcal{N}_{ijk}} \left\| \Phi^n_{ijk} - \Phi^n_{i'j'k'} \right\|^2, \tag{2}$$

with the global regularization parameter α. In the following we shall assume that D can be expressed as a function of the probability distribution of matched voxel intensities $p(a,b)$. The estimated change of $C(\Phi^n)$ from moving a single displacement vector Φ^n_{ijk} to ϕ is computed by a linear approximation. The notation $C^n_{\phi,ijk}$ denotes the value of the cost function (1) when replacing the single vector Φ^n_{ijk} by ϕ.

$$\Delta C^n_{\phi,ijk} \simeq \Delta R(\Phi^n) + \sum_{a,b} \Delta p(a,b) \frac{d}{dp(a,b)} D(\Phi^n). \tag{3}$$

The distribution $p(a,b)$ may be estimated directly from the two-dimensional discrete histogram $h^n(a,b)$ given the current estimate of the displacement field:

$$\begin{aligned} h^n(a,b) &= \sum_{\boldsymbol{v} \in \mathcal{G}} \delta\{\mathcal{A}(\boldsymbol{v}) - a, \mathcal{B}(\boldsymbol{v} + \Phi^n_{\boldsymbol{v}}) - b\} && (4) \\ p(a,b) &\simeq h^n(a,b)/N, \quad N = \sum_{a,b} h^n(a,b), && (5) \end{aligned}$$

where $\delta\{\cdot\}$ is the Kronecker delta function. A limited neighborhood centered around the previous iteration vector is searched exhaustively evaluating equation (3) at integer displacements of the corresponding block of template voxels. The neighborhood is limited to ± 2 voxels shift (subsampled MR voxels) along the 3 axes giving a total of 125 search locations for each individual vector update.

$$\begin{aligned} \Phi^{n+1}_{ijk} &= \underset{\phi \in S^n_{ijk}}{\text{minarg}} \ \Delta \hat{C}_\phi && (6) \\ S^n_{ijk} &= \left\{ \phi \in \mathcal{Z}^3 \middle| \, |\phi - \Phi^n_{ijk}|_\infty \leq 2 \right\} && (7) \end{aligned}$$

A simple estimate of the change of the probability distribution $p(a,b)$ caused by replacing Φ^n_{ijk} with ϕ can be obtained by assuming that the box of voxels associated with the vector Φ^n_{ijk} has a constant change of displacement. Let Q_{ijk} represent the set of coordinates indexing the template box of voxels associated with vector Φ_{ijk} so we can write

$$\Delta p_\phi(a,b) \simeq \frac{1}{N} \sum_{\boldsymbol{v} \in Q_{ijk}} \delta\{a - \mathcal{A}(\boldsymbol{v}), b - \mathcal{B}(\boldsymbol{v} + \phi)\} - \frac{1}{N} \sum_{\boldsymbol{v} \in Q_{ijk}} \delta\{a - \mathcal{A}(\boldsymbol{v}), b - \mathcal{B}(\boldsymbol{v} + \Phi^n_{ijk})\}$$

Using this result we get

$$\begin{aligned} \Delta D_\phi &\simeq \sum_{a,b} \Delta p_\phi(a,b) \frac{dD(\Phi^n)}{dp(a,b)} \\ &\simeq \sum_{\boldsymbol{v} \in Q_{ijk}} \frac{1}{N} \left\{ \frac{dD(\Phi^n)}{dp(\mathcal{A}(\boldsymbol{v}), \mathcal{B}(\boldsymbol{v} + \phi))} - \frac{dD(\Phi^n)}{dp(\mathcal{A}(\boldsymbol{v}), \mathcal{B}(\boldsymbol{v} + \Phi^n_{ijk}))} \right\} \\ &= c + \frac{1}{N} \sum_{\boldsymbol{v} \in Q_{ijk}} \mathcal{T}\{\mathcal{A}(\boldsymbol{v}), \mathcal{B}(\boldsymbol{v} + \phi)\}, && (8) \\ \mathcal{T}\{a,b\} &= \frac{d}{dp(a,b)} D^n_\Phi && (9) \end{aligned}$$

where c is independent of Φ^{n+1}_{ijk}. Equation (8) initially applies a first order approximation[3] to the change in ΔD_ϕ wrt. $p(a,b)$, and then the aforementioned approximation to the change $\Delta p(a,b)$. Thus, the final change in D is estimated by summing values from a look-up table $\mathcal{T}(a,b)$ which can be computed from the current estimate of the $p(a,b)$ distribution, and has a size proportional to the square of the number of discrete levels that a and b may take, i.e. 256 × 256 if 8 bit internal representation is used. This look-up strategy makes the optimization of the global similarity measure computationally feasible since the bulk of computational work is reduced to indexing and summation operations. An update of an individual vector consists of, among all $\phi \in S_{ijk}$, choosing the one that minimizes the cost function locally, i.e. minimizing

$$\begin{aligned} \Delta \hat{C}_\phi &= \frac{1}{N} \sum_{\boldsymbol{v} \in Q_{ijk}} \mathcal{T}\{\mathcal{A}(\boldsymbol{v}), \mathcal{B}(\boldsymbol{v} + \phi)\} \\ &+ \alpha \sum_{i'k'j' \in \mathcal{N}_{ijk}} \left\| \phi - \Phi^n_{i'j'k'} \right\|^2. && (10) \end{aligned}$$

This approach saves a considerable amount of computing time when the voxel similarity measure is a function of the entire ditribution $p(a,b)$, as is the case for the CC and MI similarity measures. When a single displacement vector is moved, only the sum of the elements $\mathcal{T}(a,b)$ where $p(a,b)$ changes needs to be recomputed in order to estimate ΔD. Alternatively, to compute D from $p(a,b)$ by direct evaluation for every single vector update, $256 \times 256 = 65536$ values would need to be recomputed (for 8 bit sampled images), compared to looking up $4 \times 4 \times 4 = 64$ table values using Equation (10), i.e. a factor ~ 1000 less operations.

The vectors are updated in a 3D checker-board fashion, i.e. vectors Φ_{ijk} for which $i + j + k$ is even

[3]It should be noted that the first order approximation poses a problem if D is not differentiable at $p(a,b) = 0$. We shall address this problem in Section II-B .

are updated in the first round and the remaining vectors in the second round. This is done to increase the convergence speed of the cost function (1). Equation (10) shows that a vector update has a Markov property that each update depends only on the 6 nearest vector neighbors—this means that by updating first one half of the 3D checker-board and then the other half, the second half of the updates will be experiencing the neighbors Φ^{n+1} field rather than the Φ^n field. This will in effect double the convergence rate.

Finally, the displacement field is smoothed in between iterations by applying a 3D Gaussian spatial low-pass filter. The size of the Gaussian kernel FWHM, which we will denote β, is a smoothness parameter that needs to be carefully set together with the α of Equation (2). The search algorithm chooses among discrete values of ϕ (minimizing ΔD_ϕ) when updating Φ^n_{ijk}. This makes the solution field consist of discrete values no matter what value the neighborhood regularizer α has. Direct smoothing using the 3D Gaussian reduces the discretization artifacts.

B. Similarity Measures

Three different similarity measures are described here, least squares (LS), correlation coefficient and mutual information (MI). The look-up table $\mathcal{T}(a,b)$ is computed at each iteration step as listed in Table II, see Appendix II for further details. In case of the correlation and MI measures, $p(a,b)$ must be estimated between iterations, which involves the overhead of computing $\mathcal{B}(\boldsymbol{v}+\Phi^n_{\boldsymbol{v}})$. For the LS measure, $\mathcal{T}(a,b)$ may be computed on line, i.e. with no table actually computed which makes the LS method slightly faster.

The Least Squares (LS) method using the equation in Table II is quite sensitive to global intensity variations in the MR scans. The LS algorithm referred to in the second part of the paper is actually a modified LS: The estimated change in the distance measure by replacing Φ^n_{ijk} with ϕ can be obtained by modifying Equation (8) to become

$$\Delta D_\phi \simeq c + \sum_{\boldsymbol{v}\in Q_{ijk}} \frac{1}{N}\left\{\mathcal{A}(\boldsymbol{v}) - \bar{\mathcal{A}} - \mathcal{B}(\boldsymbol{v}+\phi) + \bar{\mathcal{B}}\right\}^2 \tag{11}$$

where $\bar{\mathcal{A}} = \sum_{\boldsymbol{v}\in Q_{ijk}} \mathcal{A}(\boldsymbol{v})/N_Q$ and $\bar{\mathcal{B}} = \sum_{\boldsymbol{v}\in Q_{ijk}} \mathcal{B}(\boldsymbol{v}+\Phi^n_{ijk})/N_Q$. This makes the algorithm fit the edges rather than the local intensity level of the data.

The estimation of $p(a,b)$ using Equation (5) poses a problem in the case of MI. It is evident from Table II that $\mathcal{T}(a,b)$ breaks down in the neighborhood of $p(a,b) = 0$. Increasing the bin size of the histogram $h^n(a,b)$ so that no bin gets a zero count would solve the problem but throws away intensity resolution. Increasing the bin-size can be thought of as smoothing the histogram with a box-filter, so instead of using the same filter size all over the histogram, an adaptive smoothing is performed. Areas of $h^n(a,b)$ with low counts are smoothed with a large scale kernel (similar to a large bin), areas with higher counts with a smaller kernel. The kernel used is a Gaussian kernel with a FWHM adjusted so that the signal to noise ratio of the smoothed histogram counts is approximately constant, assuming local uniformity of $p(a,b)$. The signal to noise ratio is computed from the variance in the binomial distribution: Let $h(a,b)$ be the current histogram of matched intensities and g_f a Gaussian blurring kernel of FWHM equal to f. Assuming local uniformity of $h(a,b)$ we may estimate a local signal to noise ratio (SNR) using

$$\begin{aligned}
\mathrm{SNR} &\simeq \frac{h(a,b)}{\sqrt{\mathrm{Var}\{h * g_f(a,b)\}}} \\
&\simeq \frac{h(a,b)}{\sqrt{\sum_{a',b'} g_f^2(a',b')h(a-a',b-b')}} \\
&\simeq \sqrt{\frac{\pi h(a,b) f^2}{2\log 2}}
\end{aligned} \tag{12}$$

so that

$$f \simeq \mathrm{SNR}\cdot\sqrt{\frac{2\log 2}{\pi h(a,b)}} \tag{13}$$

In the actual implementation the histogram is smoothed with 5 blurring kernels (FWHM = 2, 3, 5, 7 and 10 histogram indices). For each (a,b), Equation (13) selects the correct local blurring size among these 5 histograms. In order to stabilize the procedure, a blurred histogram (FWHM=2) is used for $h(a,b)$ in Equation (13). A SNR value of 10.0 produced good results. After applying this scheme, no bins will contain the zero count value (because of the global range of the Gaussian kernel), and the parts of the histogram that have a high number of counts preserve their information.

C. Discussion

Both the sum of squares and the correlation similarity measures rely on the fundamental assumption that the residual difference between the warped subject image and the template is white Gaussian noise. This can be realized by thinking of the cost function (1) as a negative log posterior distribution. The regularizer R is a prior distribution on the displacement vectors while the similarity measuring term D is a model of the residual noise, which in the quadratic cases is a product of independent Gaussian distributions. We know, however, that this assumption is

TABLE II

LOOKUP-TABLE FUNCTIONS $\mathcal{T}(a,b)$ DERIVED (SEE APPENDIX II) FROM 3 DIFFERENT SIMILARITY MEASURES D. NOTE THAT CONSTANT TERMS AND SCALING FACTORS ARE IGNORED IN THE $\mathcal{T}(a,b)$'S, AND THAT THE D LISTED FOR MI IS ACTUALLY THE NEGATIVE MUTUAL INFORMATION—THE WARP ALGORITHM MINIMIZES D, THEREBY *maximizing* THE MUTUAL INFORMATION FOR $p(a,b)$. (c_1, c_2) IS THE EIGENVECTOR WITH THE SMALLEST EIGENVALUE OF THE 2×2 COVARIANCE MATRIX BETWEEN INTENSITIES a AND b.

Method	Distance measure	$\mathcal{T}(a,b)$
Least Squares	$\sum_{a,b} p(a,b)(a-b)^2$	$(a-b)^2$
Correlation Coefficient	$\rho_{ab} = \frac{\sum_{a,b} p(a,b)(a-\bar{a})(b-\bar{b})}{\sqrt{\sum_a p(a)(a-\bar{a})^2 \sum_b p(b)(b-\bar{b})^2}}$	$(c_1(a-\bar{a}) + c_2(b-\bar{b}))^2$
Mutual Information	$-\sum_{a,b} p(a,b) \log \frac{p(a,b)}{p(a)p(b)}$	$-\log \frac{p(a,b)}{p(a)p(b)}$

incorrect in the sense that the subject anatomy *is not* a deformed instance of the template anatomy—real anatomical differences certainly exist among normal brains. The major disadvantage of the squared similarity measures is that the anatomical differences that cannot be corrected for by the warp are penalized with a factor depending on how the image modality represents the mismatched tissue types as image intensities. There is no reason to believe that a mis-registration of e.g. gray matter with CSF should be penalized by a factor depending only on the difference in the intensity representations of those two tissue types. The MI on the other hand, penalizes (mis)matches with a factor depending on the resulting specificity (i.e. entropy of the conditional distribution $P(b|a)$ of subject voxel intensities b given corresponding template voxel intensity a) of the mapping between image intensities. This is certainly a more general approach, and it may be a better criterion.

Another relevant issue is the number of iterations needed for the warp to converge at each resolution step. Our experience has been that MI requires twice the number of iterations to converge (typically 8) compared to the sum of squares method (typically 4) so that the total computation time is approximately doubled. This is because the distance measure $D(a,b)$ is a more complex function of the displacement field when using MI than LS or correlation, so that more iterations are needed in each resolution step.

D. Smoothness Constraints

Several factors are working together to ensure the smoothness of the displacement field. The reason for requiring smoothness is simply a question of controlling noise in the resulting displacement map. The factors are:

- *Use of a regularizer, or prior term in the global cost function (1).* The term $R(\Phi)$ is essentially a Gaussian Markov field prior. The frequency power spectrum for this prior may be calculated analytically, see e.g. [28]. Previous work has been published applying physical/mechanical prior distributions to the deformation problem [12], [29], [30], [31], [32], [33]. However, considering that the deformation itself is non-physical (i.e. the subject anatomy *is not* a deformed instance of the template anatomy), we argue that neither does the prior distribution have to be. For arguments in favour of physical regularizers see [19], [34].
- *Controlling the sampling of the displacement field.* By limiting the number of vectors in the representation of Φ we efficiently limit the upper frequency components of the deformations that can be represented. In [35], the displacement field is expressed as an expansion of smooth basis functions thereby restricting the resolution of the resulting displacement field. In [21] a hierarchical search for the displacement field is performed by representing the displacement field in wavelet components, and gradually increasing the number of components used.
- *Smoothing of the image volumes.* This technique is used implicitly in the sub-sampling of the subject and template volumes. It is also used in [35] and in the ANIMAL algorithm described in [9], [10], [13], [36].
- *Direct smoothing of the vector field components.* This approach is a direct way of shaping the frequency characteristics of the displacement field. The parameter β controls the size of the Gaussian low-pass filter used. Low-pass filtering of the displacement field vectors is also used in [8].

The problem of actually choosing the parameters used is a topic of current investigation. At each resolution step a new set of parameters is applied. Preliminary work based on the effect of the warp on the analysis of a functional data set has indicated that α and β cannot substitute for one another, although there exists a correlation in the optimal (with respect to the resulting functional alignment as described in Section IV) setting of the two parameters. It should further be noted that the setting of α depends strongly on the voxel similarity measure used, since the effect of the regularizing term changes with the dynamics of the match term in Equation (1).

III. Data Acquisition

The functional activation study is a saccadic eye movement experiment [37] done at Rigshospitalet, Copenhagen, Denmark, consisting of 5 male right-handed subjects, each scanned 8 times (2 baseline, 6 different graduated target frequencies) during two scanning sessions, over two days with randomized ordering of scans. An array of LED's was located in an arc at angles -40° to 40°. The subjects performed visually guided voluntary anti-saccades. This paradigm involved subjects fixating on a central LED until a random target LED was lit. The task was to perform a saccade to the LED contralateral to the lit target. The target presentation frequency was varied across scans from 0.05, 0.1, 0.3, 0.5, 0.7 to 0.9 Hz. Two baseline scans were acquired with the subject fixating at the central LED. The scanner is a whole body PET-scanner (Advance, General Electric, [38]) with an image resolution of approximately 5 mm in all directions. The scans were reconstructed with 35 slices, voxel size $2.0 \times 2.0 \times 4.25$ mm covering a field of view of $25.6 \times 25.6 \times 15.2$ cm. In addition, MR scans were obtained for the 5 PET subjects plus an additional subject to act as an alignment template. The scans were T1-weighted acquisitions using a Siemens Magnetom Vision scanner located at Hvidovre Hospital, Copenhagen, with $0.98 \times 0.98 \times 1.17$ mm voxel size for two of the five subjects and $0.98 \times 0.98 \times 1.00$ mm voxel size for the remaining 4 subjects, all acquired using a 3D MPRAGE sequence. The PET experiment consisted originally of 8 subjects out of which 2 did not wish to have the MR scan, and 1 subject was removed because his MRI showed an abnormality in the cerebellum, leaving 5 subjects for this experiment.

Other workers have applied a template volume which is built as an average image over normal individuals in stereotactic space (e.g. [10]averaged template does not contain the sharp anatomical boundaries that exist in an individual MR image. The impact of the tendency to bias the results of the analysis towards the characteristics of a specific brain template remains an open research issue. However, our registration algorithm itself does not impose the use of a specific type of anatomical template.

A. Preprocessing

A common preprocessing step was performed on all scans. Each subject's PET scans were intra-subject aligned using a 6 parameter rigid-body rotation and translation (AIR, [3]) to match each scan to the subject's first scan. A mean scan was computed for each subject and a slice based[4] threshold of 45% of the slice maximum was applied to slices containing brain voxels, to generate the brain/non-brain binary masks. The injected dose/weight ratios were not available, so global scaling of each scan was achieved using a voxel based Scaled Subprofile Model (SSM) analysis (each scan is assigned a global scaling factor g_{lm}, see Section IV and [40], [41]). No spatial smoothing was applied to the scans following the reconstruction, yielding an effective resolution of approximately 5 mm in plane and 6.5 mm axially (PET images reconstructed with a Hann filter in-plane and an unappodized Colsher filter in the axial direction).

The MR scans were corrected for intensity inhomogeneity, which was predominant in the axial (Z) direction. This was accomplished by computing the mean of the top 5% intensity voxels for each slice containing brain tissue, and normalizing the slice intensities by fitting a 6'th order polynomial through these values as a function of the slice number, which gave a visually improved result. The MR scans were then carefully stripped from skull and scalp using a manual interactive stripping/region drawing tool, retaining mainly CSF, white matter and grey matter within a contiguous mask.

B. Alignment Schemes Compared

Five registration schemes, which we will name A–E, were set up for comparison as illustrated in Figure 1. Scheme A was our standard alignment procedure using a 12-parameter affine registration to align the mean PET images for each subject with a Talairach PET template volume obtained by simulating a PET scan from a high-resolution segmented MRI scan[5] [1]. In Scheme B the 12-parameter inter-subject registration was computed using each subject's stripped MR volumes and a Talairach transformed MR template brain. The PET images were aligned via this transformation, so that each PET scan was registered by

[4] A slice-based threshold was applied because this corresponds to our [1], [39]

[5] Regrettably, this MR scan was not identical to the MR template used for the other alignment schemes. However, this seems to have had no detectable impact since the alignment Schemes A & B (see Figure 1) produce very similar results.

combining the individual 6 parameter PET-PET and PET-MR transformations, and 12-parameter affine transformations in a single resampling operation. In Schemes C–E, a 3D non-linear registration field was computed on each of the subject's 12-parameter registered, stripped MR scans to match the stripped template scan. Each PET scan was resampled combining the non-linear warp, the individual 6 parameter PET-PET, PET-MR, and 12-parameter MR to template transformations in a single resampling operation. The different non-linear registration Schemes C-E differ only in the voxel similarity measure used, as described later.

In all three cases, both the raw PET images and the binary masks (1 is brain tissue, 0 is background) generated from the mean PET images were resampled using tri-linear interpolation in the same resolution as the original data ($2 \times 2 \times 4.25\,\mathrm{mm}$). New binary subject masks were formed by thresholding the resampled masks at 0.5. All 5 resampled masks were then logically AND'ed to find the resulting global mask used for the functional analysis.

Apart from the absence of a step with spatial smoothing of the PET images, Scheme A was identical to our standard procedure used for group analysis of PET activation studies also described in [1], [39]. This is a relevant comparison since in many cases, MRI scans are not acquired routinely with the PET scans.

One may argue that method B–E are more sensitive to errors in the 6 parameter registration than method A since they concatenate two rigid-body transformations while method A only involves one. This suggests the alternative procedure of aligning each of the subject's PET scans to the MR scan individually. That approach has not been used because it is our experience that the noise in the raw PET scans gives problems with the cross modality registration, so that the total registration error becomes larger. A possible solution to this dilemma for future investigation is offered in [4] by reconciling all possible pairwise rigid-body transformations using an iterative optimization scheme.

For comparison, three similarity measures for the non-linear warp were selected: The modified least square (LS), correlation coefficient (CC) and mutual information (MI), producing three separate nonlinearly aligned data sets (which we shall refer to as C, D and E, respectively). At this point we face the problem of setting the regularization parameters α Equation (10) and β of Section II-A . Three resolution steps were used ($64 \times 64 \times 40$, $128 \times 128 \times 80$, and $256 \times 256 \times 160$, at all times having one displacement vector for each $4 \times 4 \times 4$ block of voxels yielding a displacement field of increasing resolution, with approximately 4 mm at the highest step) and the regularization parameter α was set to 150, 35 and 6 for the three resolution steps for the warp using LS as well as CC, and to 0.275, 0.066 and 0.011 for the warp using MI. These values were chosen so that the average variation of the regularization term in Equation (10) achieves the same order of magnitude as the average variation of the similarity term during the warp iterations. The filtering parameter β which controls the Gaussian FWHM (measured in vector indices) of the direct smoothing of the vector field between iterations was chosen (by trial and error and visual inspection of the generated displacement fields) as 0.65, 0.41 and 0.38 for the three resolution steps, or 2.60, 1.64 and 1.52 image voxels, respectively.

IV. Data Analysis

A. *Multivariate Functional Analysis*

Schemes A, B and the three different non-linear registration Schemes C, D and E produced five sets of stereo-tactically registered PET scans. The scans of each data set were passed through an SSM (Scaled Subprofile Model) analysis (see [40], [41] for a description of how to perform the actual computation procedure) and a subsequent Canonical Variate Analysis (CVA, see appendix or e.g. [42, Ch. 12]) of the resulting eigenvectors.

The SSM model is based on an orthogonal decomposition of the covariance structure of the log-transformed data, with a global scaling factor g_{lm} for each scan. Let v_{ilm} denote the detected $\left[{}^{15}\mathrm{O}\right]$water concentration at each voxel location i and scan indexed by subject $m = 1 \ldots M$ and session scan number $l = 1 \ldots L$. With this notation, the SSM model is

$$
\begin{aligned}
v_{ilm} &= g_{lm}\left(r_i + i_{ilm}\right), \qquad \sum_{lm} i_{ilm} = 0 \quad (14) \\
i_{ilm} &= \sum_{k=1}^{K} h_{ik}\sqrt{\lambda_k}\, s_{klm} = \sum_{k=1}^{K} h_{ik}\mathrm{SSF_{lmk}}, \quad (15)
\end{aligned}
$$

where K is the number of orthogonal components in the model. In the analysis used in the following, K equals the estimated number of degrees of freedom in the data matrix. The vector r_i is the mean activation image, and the i_{ilm} matrix contains the interaction terms. This interaction matrix is decomposed using the orthonormal vectors h_k each of which describe a covariance relationship between the voxels. The eigen-vectors SSF_{lmk} have corresponding eigenvalues λ_k, sorted so that $\lambda_1 > \lambda_2 > \ldots > \lambda_K$, which describe the amount of variance accounted for by the

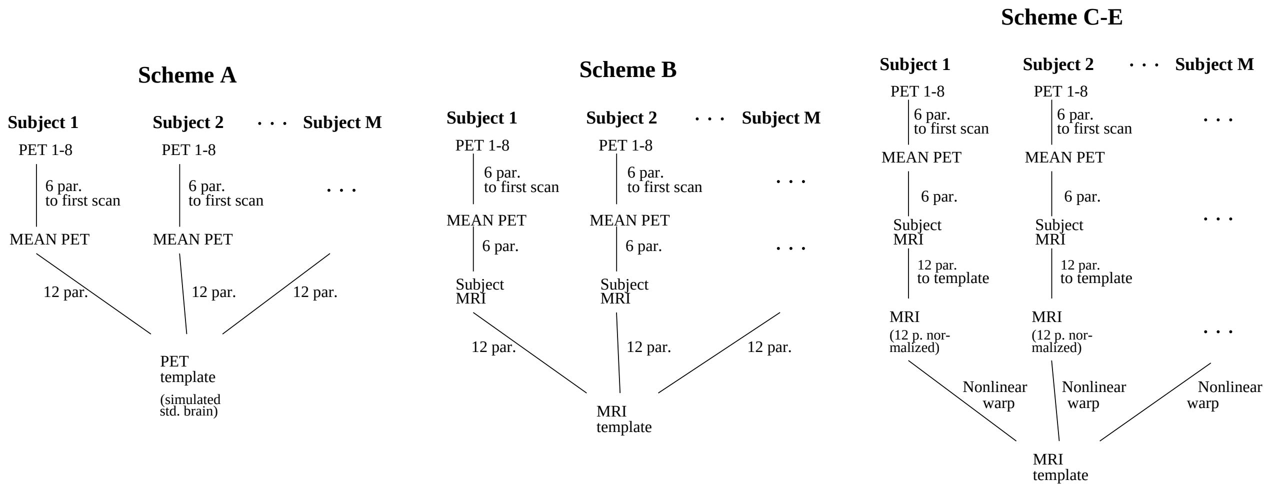

Fig. 1. The alignment schemes used for co-registering PET images across subjects. Scheme A is affine registration via PET, Scheme B is affine registration via MRI, while Schemes C–E use a non-linear method with varying voxel similarity measures applied.

k'th eigen-vector[6].

The SSM analysis is purely data-driven, which means that no information about the particular activation (paradigm) state goes into the model. Nevertheless, in multisubject activation experiments it is commonly the case that the first M-1 eigenvectors (M equals the number of subjects) describe variation between individuals and the following eigenvectors describe variation related to the activation experiment, see e.g. [44], [45]. This division of signal variance into an inter- and an intra-subject component is also visible in Figure 5 from the drop in the magnitude between eigenvalues 4 and 5 and was also found in the original group of 8 subject's PET scans [46].

For the particular experiment there were 6 different saccadic frequencies and 1 fixation condition. The variation between these 7 groups may be analyzed within the space spanned by the SSM eigenvectors using a Canonical Variate Analysis (CVA) [42, Ch. 12] to find a low dimensional subspace containing all the experimental structure as previously utilized in [47], [48]. The CVA analysis generates an orthogonal decomposition of the between group variance, so that the n'th canonical vector holds the n'th largest ratio of between group covariance and within group covariance. The CVA is performed only in the subspace spanned by the 12 largest SSF eigenvectors. The eigenvalues of the CVA analysis describe the distribution of the total variance into the respective canonical vectors. The computational details are outlined in the appendix.

B. Univariate Variances

More direct univariate variance measures were also computed. Let subject $m = 1 \ldots M$ and session scan number $l = 1 \ldots L$ ($Q = L \cdot M$). The total inter- and intra-subject variances normalized by the number of voxels retained, I, are computed from the interaction i_{ilm}:

$$\begin{aligned}
\sigma^2_{\mathrm{TOTAL}} &= \frac{1}{I(L \cdot M - 1)} \sum_{i,l,m} (i_{ilm} - i_{i\cdot\cdot})^2 \\
&= \frac{1}{I(L \cdot M - 1)} \sum_{i,l,m} i^2_{ilm} \\
\sigma^2_{\mathrm{INTRA}} &= \frac{1}{I(L \cdot M - M)} \sum_{i,l,m} (i_{ilm} - i_{i\cdot m})^2 \qquad (16) \\
\sigma^2_{\mathrm{INTER}} &= \frac{1}{I(M - 1)} \sum_{i,m} (i_{i\cdot m} - i_{i\cdot\cdot})^2 \\
&= \frac{1}{I(M - 1)} \sum_{i,m} i^2_{i\cdot m}
\end{aligned}$$

C. Anatomical Analysis

The inter-subject variance may also be computed for the MR images (see Table III). Figure 2 shows selected orthogonal slices of the template MR, and the registered MR scans for each of the results (averaged across the five subjects) of 12-parameter alignment, and the three warps. The left column is the template volume, and the four other columns are the averaged aligned MR volumes using the four methods B–E.

[6]Let scan l, subject m be indexed by a single index q. Now, if $i_{iq} = \sum_{k=1}^{K} \mathrm{SSF_{qk}h_{ik}}$ then the matrices h_{ik} and the SSF_{qk} (Scan Scaling Factors) are in essence the components of a Singular Value Decomposition (SVD, see e.g. [43] although they are not computed directly as such because of the g_q (g_{lm}) factors, see [40]) of the interaction matrix i_{iq}, in matrix notation: $i_{iq} = (\mathbf{USV}^T)_{iq}$ with $\mathbf{U}_{ik} = h_{ik}$ and $(\mathbf{SV}^T)_{qk} = \mathrm{SSF}_{qk}$.

The non-linear warp produces images that vary the least from the template volume, so that the resulting mean images are sharpest, compared to the affine transformation.

The non-linear registrations resulted in a better alignment of the gross shape of the brain. A visual inspection of the warped volumes indicated that MI performed slightly better than CC and LS, but none of the registration methods were able to perfectly match the cerebral sulci.

D. Functional Analysis Results

Numerical results from our analysis are summarized in Table III. Note that the non-linear registration retains more voxels in the analysis, the warp using MI retained 2.5% more voxels than the 12-parameter method. This occurs because the non-linear registration fits the gross shape of the brain quite well. We only retain voxels that are present in all scans, and this makes the analysis quite sensitive to registration errors at the edges of the masks—in this way the number of voxels becomes a measure of registration performance, and the non-linear registration is seen to best preserve the field of view of the analysis. Since most of the extra voxels in the non-linear registration analysis reside in grey matter areas of the cortex, we may be retaining important information about cortical activation that would be excluded using the 12-parameter alignment. Note that the number of voxels for Scheme A (PET-to-PET registration) is not directly comparable to the other schemes since the PET template brain was smaller than the MR template used in Schemes B and C–E. Figure 3 shows horizontal slices (+45 mm above and 10 mm below AC-PC plane) of the first canonical vector, i.e. the direction in scan space of largest functional discrimination between the groups representing saccadic frequencies. It is evident that the non-linear registration (the three right-most images) is better at preserving edge voxels in the analysis. The upper slice contains an activation focus in the dorso-lateral prefrontal cortex, which was undetected by the 12-parameter affine registration scheme B because the voxels near the edge of the brain were masked out. The major activation foci in the visual cortex were captured by all the registration schemes tested.

The change in total, inter- and intra-subject variance is shown graphically in Figure 4: The drop in total variance ($2.09/2.06 \cdot 10^{-2}$ affine A/B to $1.85/1.76/1.75 \cdot 10^{-2}$ non-linear C/D/E) is seen to arise mainly from a change in the inter-subject variance; the intra-subject variance is virtually unaffected by the alignment scheme. The inter-subject variance drop is absorbed mainly by the first 4 eigenvectors of the SSM analysis, as can be seen in Figure 5. The plot shows the SSM eigen-spectra (λ_k) with each eigenvalue normalized by the number of voxels retained in the analysis. Also, the sum of the first 4 eigenvalues listed in Table III reflects the change in inter-subject variance σ^2_{INTER}.

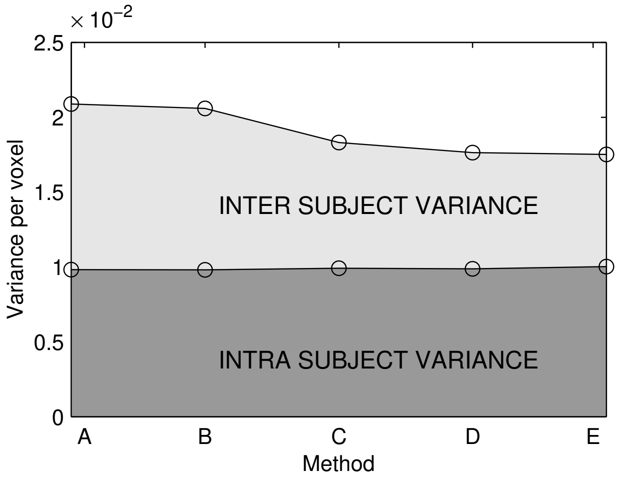

Fig. 4. Division of the total functional variance into inter- and intra subject contributions. A: Alignment using PET to PET 12-parameter affine; B: 12-parameter affine via MRI; C–E: Non-linear warp using Least Squares (LS), Correlation Coefficient (CC) and Mutual Informaion (MI) respectively as similarity measures.

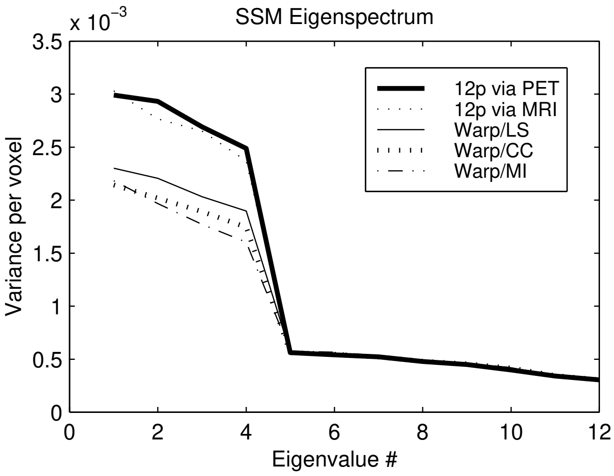

Fig. 5. The plot shows the SSM eigenspectrum obtained from each of the 2 affine and 3 non-linear alignment experiments: Least Squares (LS), Correlation Coefficient (CC) and Mutual Informaion (MI). The SSM eigenvalues correspond to the eigenvalues of the covariance matrix for the i_{ilm} vectors of Equation (15).

These results indicate that Scheme E (warp using MI) may be the superior registration method. But until now, we have been looking only at very direct

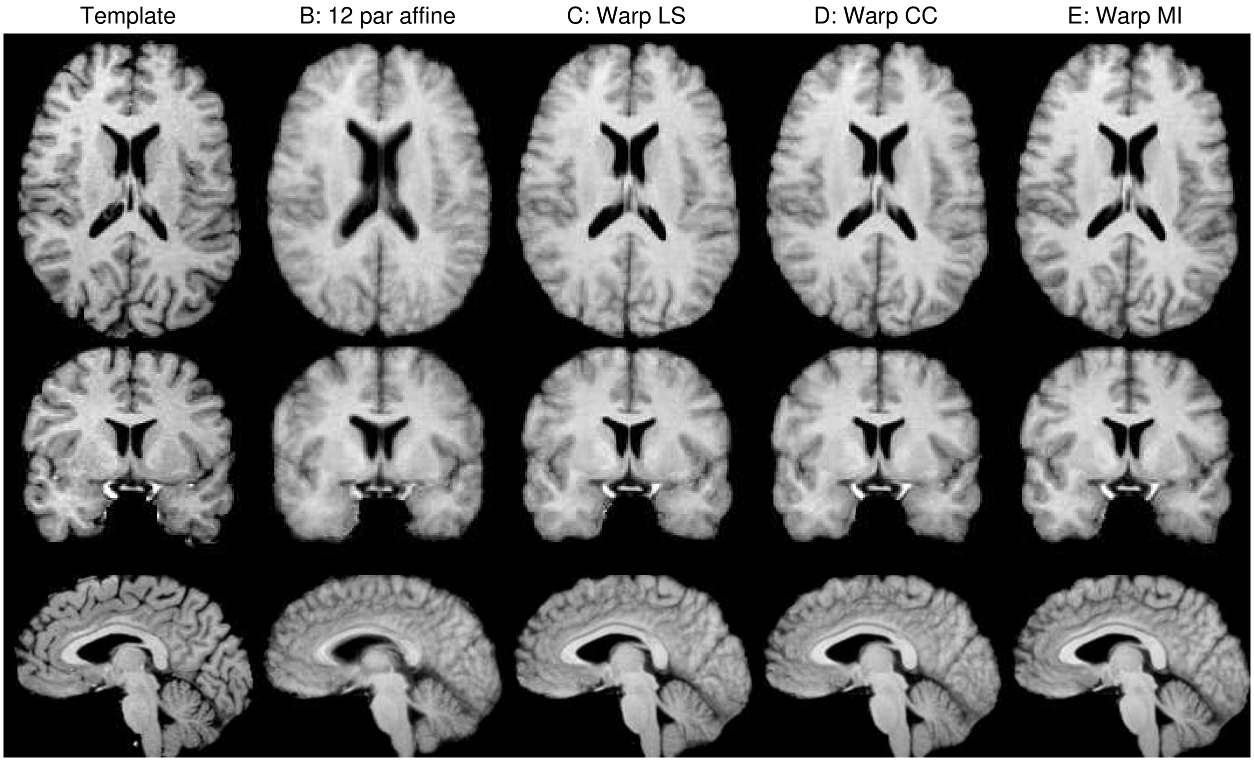

Fig. 2. Comparison of MR alignments. Images in the leftmost column are from the template volume. Images in the four other columns are mean slices across aligned subjects for each of the four alignment methods (B–E): 12-parameter affine and warps using the LS, CC and MI similarity measures, respectively. The sharper MI images indicate better alignment by this method.

variance measures, and we have seen that Scheme E has the largest impact on the functional alignment and reduces the inter-subject variance the most. This way of evaluating the registration does not measure how good a *functional* alignment we have achieved, i.e. how good is the signal to noise ratio of the observed brain signal. In the following we shall investigate how well the brain state is reflected in the scans, assuming that the optimal alignment is the one that is able to extract and separate the brain states most clearly. A ideal approach would be to measure the reproducibility or generalization ability of the spatial activation patterns obtained, but this would require a data set with many more than 5 subjects as investigated here (e.g. [49], [50]).

The canonical variate analysis gives information about how the frequency of the saccadadic eye movement affects the scans in a low-dimensional subspace embedded in the high-dimensional space spanned by all scans. All scans are grouped according to the saccadic eye movement frequency (with 7 groups in all). The canonical vectors form an orthonormal basis that maximizes the separation between these group means while minimizing the spread within groups. If the relationship between the scan and the saccadic frequency is purely linear, the variation between the 7 frequencies can be described adequately with a linear model. Figure 6 shows a plot of the canonical projections for all 40 scans registered using Scheme B. The circles indicate 90% confidence areas for the location of each saccadic frequency group mean [42, Ch. 12]. Although mild non-linearities in the saccadic functional activation patterns from this experiment have been reported [46], Figure 6 demonstrates that the majority of the group structure is well described by the first canonical component. We therefore assume that the remaining canonical components (λ_2–λ_6) describe 'noise' in the activation, i.e. patterns not directly related to the paradigm. Thus, the ratio of the first canonical component to the sum of the rest ("Signal to Noise Ratio") is an indication of how well the brain activation signal is captured. As shown in Table III it is actually Scheme C with a ratio of 7.7 which performs best in this sense, while Scheme E with the lower ratio of 7.3 produced the lowest inter-subject variance. We also see that the non-linear registration methods tested all outperform the 12-parameter affine registration of Schemes A and B. Figure 7 shows a scatter plot of the first versus second canonical eigenvalue from the CVA analysis for

TABLE III

Comparison of effect of the warp vs. affine transformations on selected variance measures per voxel from the anatomical (MRI) and functional (PET) analysis.

Method	A	B	WARPS: C–E		
	12p PET-PET	12p via MRI	C: Warp/LS	D: Warp/Corr	E: Warp/MI
[a]MRI σ_{INTER}	-	30.5	26.6	25.4	24.8
No. PET voxels retained, I	67621[b]	81689	83349	82908	83773
$\sigma^2_{\mathrm{TOTAL}}$	$2.089{\cdot}10^{-2}$	$2.060{\cdot}10^{-2}$	$1.832{\cdot}10^{-2}$	$1.764{\cdot}10^{-2}$	$1.752{\cdot}10^{-2}$
$\sigma^2_{\mathrm{INTER}}$	$1.349{\cdot}10^{-2}$	$1.315{\cdot}10^{-2}$	$1.024{\cdot}10^{-2}$	$0.946{\cdot}10^{-2}$	$0.913{\cdot}10^{-2}$
$\sigma^2_{\mathrm{INTRA}}$	$1.094{\cdot}10^{-2}$	$1.093{\cdot}10^{-2}$	$1.105{\cdot}10^{-2}$	$1.100{\cdot}10^{-2}$	$1.117{\cdot}10^{-2}$
SSM[c]: $\sum_{i=1}^{4}\lambda_i/N$	$1.108{\cdot}10^{-2}$	$1.082{\cdot}10^{-2}$	$0.844{\cdot}10^{-2}$	$0.780{\cdot}10^{-2}$	$0.753{\cdot}10^{-2}$
[d]CVA: $\lambda_1/\sum_{i=2}^{6}\lambda_i$	5.505	5.803	7.653	7.397	7.307

[a]MRI σ_{INTER} is measured per voxel inside the brain mask on the inhomogeneity corrected scans. This measure is sensitive to normalization: Prior to computing MRI σ_{INTER}, each scan was normalized using the mean of manually defined samples of white matter to match the intensity of the template's white matter.

[b]The number of voxels retained in the PET-PET analysis (A) is not comparable to the MRI registrations because the PET template was smaller than the MRI template used in Schemes B and C–E.

[c]The SSM (Scaled Subprofile Model) eigenvalues λ_i describe the variance accounted for by the i'th eigenvector of the $I \times I$ covariance matrix of the i_{ilm} terms of Equation (14). In a study of M subjects, the $M-1$ (i.e.: 4) first eigenvalues usually describe inter-subject variation.

[d]The CVA eigen-values describe the dimensionality of the activation signal embedded in the high-dimensional scan space. A high value of the above λ-ratio indicates that the activation signal is adequately described by a single linear component.

each of the 5 alignment Schemes A–E.

These results seem to indicate that it is possible to 'over-warp' the data: Instead of registering functionally equivalent areas among subjects, the more detailed registration (e.g. Scheme E) applies more active deformations which *degrade* the functional registration. To illustrate this, we compute the Jacobian of the non-linear transformations. Histograms of Jacobian values (excluding background voxels) averaged over the 5 subjects are shown in Figure 8. The transformation computed by MI involves more compression ($J < 1$) and expansion ($J > 1$) of voxel elements than the transformation generated by CC and LS.

The above discussion thus leads us to conclude that although the warp of Scheme E does the best job at reducing inter-subject variation, we seem to be better off applying a registration method which does not go into the same level of detail. Presumably this is an issue of choice of regularization parameters for the warp, rather than an issue of selecting among similarity measures, but it is our experience that the MI similarity measure offers the possibility of performing registration at a higher level of detail than can be accomplished with any choice of regularization parameters for the CC and LS measures.

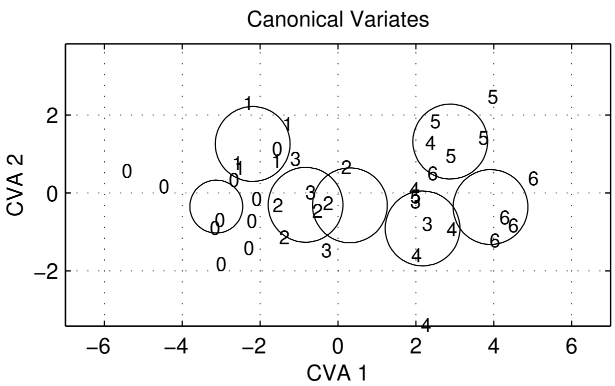

Fig. 6. Plot of first versus second canonical vectors of a Canonical Variates Analysis (CVA) of the 7 saccadic eye movement groups: 0 fixation or 0 Hz; 1: 0.05 Hz, 2: 0.1 Hz, 3: 0.3 Hz, 4: 0.5 Hz, 5: 0.7 Hz and 6: 0.9 Hz. The circles are rough 90% confidence areas for the group means. The analysis was run using the 12-parameter affine registration via MRI (Scheme B).

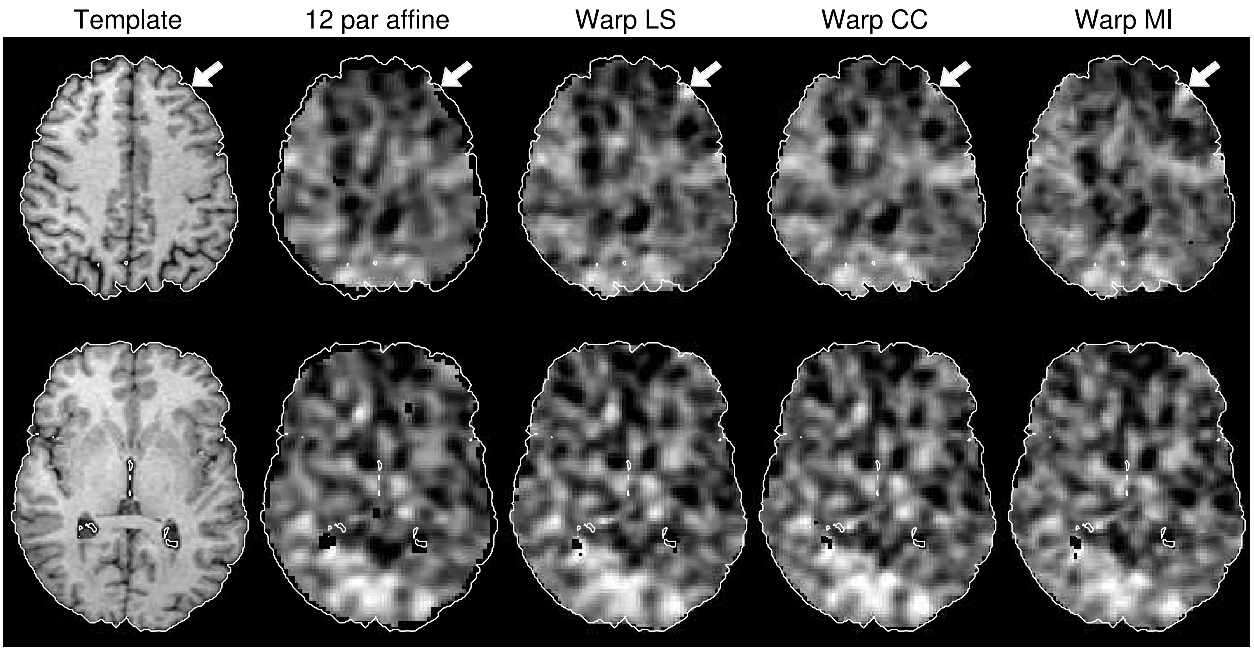

Fig. 3. Comparison of functional activation patterns. Leftmost image column shows horizontal slices (+45 mm above AC-PC plane and −10 mm below) of the template volume. The four other columns show corresponding slices of the eigen images belonging to the first canonical component, as obtained from the SSM/CVA analysis of aligned subjects for the four alignment methods : 12-parameter affine and warps using the LS, CC and MI similarity measures respectively. Note that the reduced number of voxels for the affine alignment results in a smaller field of view (only voxels present in all the subject masks enter the analysis), which causes the hot spot in the dorso-lateral prefrontal cortex to be overlooked.

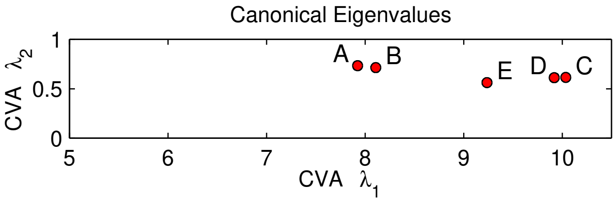

Fig. 7. Scatter plot of first versus second canonical eigenvalues of a canonical variates analysis of the 7 saccadic eye movement groups, 0 Hz to 0.9 Hz. The non-linear registration methods (C, D& E) achieve a larger 1st eigenvalue indicating better alignment. A: Alignment using PET to PET 12-parameter affine; B: 12-parameter affine via MRI; C–E: Non-linear warp using Least Squares (LS), Correlation Coefficient (CC) and Mutual Informaion (MI) respectively as similarity measures.

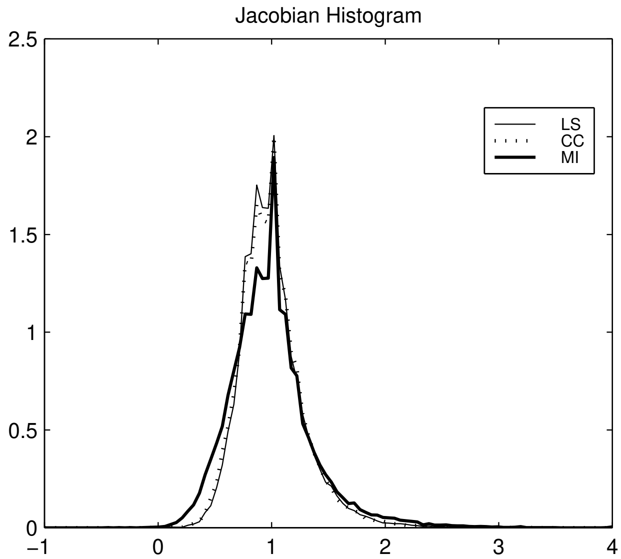

Fig. 8. Histogram of Jacobian values for the transformations computed by MI and LS. Background voxels are left out. Notice that the MI transformation has more compression ($J < 1$) and expansion ($J > 1$) of voxels than both CC and LS indicating a more active deformation.

V. Conclusion

We have presented a new method for performing fast non-linear registration of MR images. A lookup-table scheme has been presented and demonstrated to be an efficient tool for the implementation and testing of different voxel similarity measures using a model based on the joint probability distribution of matched voxel intensities.

The method has been tested on a functional $[^{15}O]$water PET data set and demonstrated to produce a registration superior to a conventional 12-parameter affine registration method. One important improvement was the enhanced registration at the edge of the cortex. This effect is expected to be even more pronounced as the number of subjects in the study increases. Furthermore, a canonical analysis of the functional activation showed that noise was removed from the data: For the non-linear registration methods tested a larger estimated functional "signal-to-noise ratio" was measured: 7.65 for the warp using LS (Scheme C) compared to 5.80 for the 12-parameter normalized data (Scheme B). The experiment described here regrettably does not contain enough subjects to produce reliable error bars on these results—more than 5 subjects are needed to get a proper statistic of the inter-subject variation. Whether or not one can expect similar improvements in functional signal space by applying the registration techniques described here on other datasets presumably depends on factors such as image resolution and noise but also more importantly on the specific activation paradigm.

We have also demonstrated that there was no advantage in using matched MRI scans with the simpler affine registration procedure. If the paradigm in question generates activation in cortical areas that are already handled quite well by the simpler registration methods (i.e. the size of the activated area is large enough to accomodate the registration errors) there may be little advantage in introducing the overhead of non-linear registration techniques. The degree to which the functional signal improves by the use of the non-linear registration thus depends on the scale of the residual anatomical variability compared to the functional variability, the size of the activated foci, and the question being asked, e.g., location versus size of activation foci.

VI. Acknowledgments

This work was supported partly by the Human Brain Project grant P20 MH57180, the Danish Research councils for the Natural and Technical Sciences through the Danish Computational Neural Network Center (CONNECT), the SLOAN Center for Theoretical Neurobiology at the SALK institute CA, and Knud Højgaards fond, Denmark. Our thanks to the anonymous reviewers for valuable comments that considerably improved the paper. The authors would further like to thank David Rottenberg, Terry Sejnowski, Alan Evans, Chin-Tu Chen and Yali Amit for valuable support and discussions.

Appendix

I. Canonical Variate Analysis

The canonical variate analysis stems from the multivariate analysis of variance and may be thought of as the analogue of PCA for grouped data. Consider an experiment with J conditions (or groups) and n_j, $j = 1 \dots J$ samples of the distribution

$$\boldsymbol{x}_{ij} \in N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}), \qquad i = 1 \dots n_j,\ j = 1 \dots J \quad (17)$$

so that each $\boldsymbol{x}_{ij}$ is a vector of size $p \times 1$. One may test if the $\boldsymbol{\mu}_j$'s are different, i.e. whether the hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_J$ can be rejected. This is done using a standard likelihood ratio test, see e.g. [42, Ch. 12]. If we are using the CVA to analyze functional brain activation in the saccadic eye movement experiment, we can group the scans according to the paradigm, in this case the saccadic frequency. Each vector $\boldsymbol{x}_{ij}$ represents[7] a brain scan with LED presentation frequency indexed by j and subject and repetition index i. Usually H_0 will be rejected (unless the paradigm did not induce significant brain activation) so we may ask more specifically about the dimensionality of the space spanned by the $\boldsymbol{\mu}$'s. Note that this is different than the univariate ANOVA case where the dimensionality is either 0 or 1. The canonical analysis performs an orthogonal decomposition of this space, so that we can inspect and test the dimensionality. Let

$$\begin{aligned}
\mathbf{W} &= \sum_{i,j} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_j)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_j)^T, \\
\mathbf{T} &= \sum_{i,j} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}})^T, \text{ and} \\
\mathbf{B} &= \sum_{i,j} n_j (\bar{\boldsymbol{x}}_j - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_j - \bar{\boldsymbol{x}})^T \qquad (18)
\end{aligned}$$

where $\bar{\boldsymbol{x}}_j$ is the mean of the group that scan number j belongs to and $\bar{\boldsymbol{x}}$ is the global mean, so that $\mathbf{T} = \mathbf{W} + \mathbf{B}$. The canonical analysis proceeds as the eigenvector analysis of the matrix $\mathbf{W}^{-1}\mathbf{B}$. The r'th canonical vector is the r'th eigenvector of this matrix and describes the linear function which obtains the largest separation of the J groups under the condition that the vector is orthogonal to the

[7] Traditionally, the $\boldsymbol{x}_{ij}$ would represent an entire scan, but this causes problems in the canonical analysis since the within group covariance matrix $\mathbf{W}$ (which is a $p \times p$ matrix) becomes singular when $p > Q - J - 1$, where Q is the total no. scans. A way to circumvent this problem is to use the p largest SSM eigenvectors to represent the scans in a lower dimension. In other words, x_{ij} represents the p values of the i'th scan in the j'th group for the p SSM eigenvectors with the largest eigenvalues. The dimensionality p should be chosen so that i) it is large enough to embed the expected variation introduced by the paradigm ii) smaller than $Q - J - 1$. In the experiment of saccadic eye movements described in this paper, p was 12. The number p should not be chosen too small since we do not know which principal components carry the signal of interest, although we expect the first $M - 1$ components to be driven mainly by inter-subject effects; the following components usually carry the activation signal of interest.

$r-1$ previous canonical directions. Note that the first canonical variate is equivalent to the Fisher's linear discriminant for optimal linear discrimination between groups. The CVA analysis can be understood as the analogue of principal components analysis for grouped data. A scatter plot of the 1st versus 2nd canonical coordinates as shown in Figure 6 can give a useful picture of the structure in the data. In such a plot, the axes are in units of standard deviations of a single data point, and we get an indication of the strength of the separation of the individual groups.

The canonical eigenvalues are central for understanding the signal-to-noise measure applied in the paper. The sum of the λ-values is the total "variance" of the group means in the transformed space where the within-group covariance matrix is unity, i.e. when representing all variables in graphs like Figure 6. One can think of the λ values as, in principle, the eigenvalues of a PCA of the group means in a coordinate system transformed so as to separate the groups the most. The distribution of λ-values (the canonical spectrum) tells us about the dimensionality of the signal space: E.g. a single large canonical value followed by small values indicates that the group means are embedded in a 1 dimensional subspace: the group means are mostly spread out along a single direction in signal space. Having two large values would suggest that the signal space has a more complex structure, two linear functions are needed to describe the data.

II. Derivation of the Look-up Table for Warp Similarity

This section describes the derivation of the voxel similarity look-up table functions listed in Table II. The look-up table functions are used to estimate the change of the global similarity measure arising from a small change in the histogram of matched voxel intensities. In the following it is assumed that the histogram is converted into a probability function by normalizing the discrete histogram $p(a,b) = h(a,b)/\Sigma_{a,b} h(a,b)$.

A. Mutual Information

Mutual information is defined by

$$D_{\mathrm{MI}}[p(a,b)] = \sum_{a,b} p(a,b) \log \frac{p(a,b)}{p(a)p(b)} \tag{19}$$

The lookup-table function is defined as $dD_{\mathrm{MI}}[p(a,b)]/\ dp(a',b')$ with the constraint that $\sum_{a',b'} dp(a',b') = 0$ which is found by straight-forward algebra:

$$\begin{aligned}
\frac{dD_{\mathrm{MI}}[p(a,b)]}{dp(a',b')} &= \sum_{a,b} \Big[\delta(a-a',b-b') \log \frac{p(a,b)}{p(a)p(b)} \\
&\quad + p(a,b) \frac{d}{dp(a',b')} \log \frac{p(a,b)}{p(a)p(b)} \Big] \\
&= \log \frac{p(a',b')}{p(a')p(b')} + \sum_{a,b} p(a,b) \Big(\frac{\delta(a-a',b-b')}{p(a',b')} \\
&\quad - \frac{\delta(a-a')}{p(a')} - \frac{\delta(b-b')}{p(b')} \Big) \\
&= \log \frac{p(a',b')}{p(a')p(b')} \qquad (20) \\
&\quad + \Big(\frac{p(a',b')}{p(a',b')} - \frac{\sum_{a,b} p(a',b)}{p(a')} - \frac{\sum_{a,b} p(a,b')}{p(b')} \Big) \\
&= \log \frac{p(a',b')}{p(a')p(b')} + \Big(1 - \frac{N_a p(a')}{p(a')} - \frac{N_b p(b')}{p(b')} \Big) \\
&= \log \frac{p(a',b')}{p(a')p(b')} + \mathrm{const.}
\end{aligned}$$

which means that a finite change $\Delta h(a,b)$ in the observed histogram results in the estimated change in the mutual information

$$\Delta D_{\mathrm{MI}} \simeq \sum_{a,b} \Delta h(a,b) \log \frac{p(a,b)}{p(a)p(b)}. \tag{21}$$

B. Entropy

Entropy is an unstable voxel similarity measure since it prefers to map all intensities to the same level (e.g. background) rather than assigning individual intensities in the template to separate intensities in the subject volume. Thus, it has not been used for any experiments in this paper. The derivation of the lookup-table function is simple

$$D_{\mathrm{EN}}[p(a,b)] = \sum_{a,b} -p(a,b) \log p(a,b) \tag{22}$$

$$\frac{dD_{\mathrm{EN}}[p(a,b)]}{dp(a',b')} = -1 - \log p(a',b') \tag{23}$$

A finite change $\Delta h(a,b)$ in the observed histogram yield the estimated change of similarity

$$\Delta D_{\mathrm{EN}} \simeq -\sum_{a,b} \Delta h(a,b) \log p(a,b). \tag{24}$$

C. Correlation Coefficient

Maximization of the correlation coefficient is achieved by gathering all samples in the histogram $h(a,b)$ along a single line. Instead of minimizing the correlation coefficient directly (which leads to a very complex expression) we will adopt a look-up table which behaves dynamically like the squared error similarity measures. This has the advantage that the warp parameter α gets approximately the same parameter value as with the squared error functions, which aids the problem of determining appropriate parameter values.

The lookup-table function reads $\mathcal{T}(a,b) = (c_1(a-\bar{a}) + c_2(b-\bar{b}))^2$ where (c_1, c_2) is the smallest eigenvector in the 2×2 covariance matrix of the matched histogram, see Equation (4). The lookup-table function measures the total variance along the direction of the smallest eigenvector, i.e. orthogonal to the direction of the linear mapping between a and b. By minimizing this variance, the correlation coefficient will be maximized. When the smallest eigenvalue is 0, the correlation coefficient will be 1.

$$\Delta D_{\text{CORR}} \simeq \sum_{a,b} \Delta h(a,b)(c_1(a-\bar{a}) + c_2(b-\bar{b}))^2 \quad (25)$$

References

[1] S. C. Strother, J. R. Anderson, X. L. Xu XL, J. S. Liow, D. C. Bonar, and D. A. Rottenberg, "Quantitative comparisons of image registration techniques based on high-resolution mri of the brain.," *Journal of Computer Assisted Tomograpy*, vol. 18, pp. 954–962, 1994.

[2] R. P. Woods and S. R. Cherry, "Mri-pet registration with automated algorithm.," *Journal of Computer Assisted Tomography*, vol. 17(4), pp. 536–546, 1993.

[3] R. P. Woods, S. R. Cherry, and J. C. Mazziotta, "Rapid automated algorithm for aligning and reslicing pet images.," *Journal of Computer Assisted Tomography*, vol. 16, pp. 620–633, 1992.

[4] R. P. Woods, Scott T. Grafton, Colin J. Holmes, Simon R. Cherry, and John C. Mazziotta, "Automated image registration: I. general methods and intrasubject, intramodality validation," *Journal of Computer Assisted Tomography*, vol. 22(1), pp. 139–152, Jan. 1998.

[5] R. P. Woods, Scott T. Grafton, John D.G. Watson, Nancy L. Sicotte, and John C. Mazziotta, "Ii. intersubject validation of linear and nonlinear models," *Journal of Computer Assisted Tomography*, vol. 22(1), pp. 153–165, Jan. 1998.

[6] J. Talairach and P. Tournoux, *Co-planar Stereotactic Atlas of the Human Brain: 3-Dimensional Proportional System. An approach to Cerebral Imaging*, Georg Thieme Verlag, 1988.

[7] D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice Hall, 1982.

[8] Y. Kosugi, M. Sase, H. Kuwatani, N. Kinoshita, T. Momose, J. Nishikawa, and T. Watanabe, "Neural network mapping for nonlinear stereotactic normalization of brain mr images," *Journal of Computer Assisted Tomography*, vol. 17(3), pp. 455–460, 1993.

[9] D. L. Collins, P. Neelin, T. M Peters, and A. C. Evans, "Automatic 3d intersubject registration of mr volumetric data in standardized talairach space," *Journal of Computer Assisted Tomography*, vol. 18(2), pp. 192–205, 1994.

[10] D. L. Collins, T. M Peters, and A. C. Evans, "An automated 3d non-linear image deformation procedure for determination of gross morphometric variability in human brain," *Proceedings of Conference on Visualization in Biomedical Computing*, vol. SPIE Vol. 2359, pp. 180–190, 1994.

[11] M. E. Alexander and R. L. Somorjai, "The registration of mr images using multiscale robust methods," *Magnetic Resonance Imaging*, vol. 14, no. 5, pp. 453–468, 1996.

[12] R. Bajcsy and S. Kovacic, "Multiresolution elastic matching," *Computer Vision and Graphics Image Processing*, vol. 46, pp. 1–21, 1989.

[13] A. C. Evans, D. L. Collins, and C. J. Holmes, "Automatic 3d regional mri segmentation and statistical probability anatomy maps," *In press*, 1995.

[14] C. Davatzikos, "Spatial normalization of 3d brain images using deformable models," *Journal of Computer Assisted Tomography*, vol. 20(4), pp. 656–665, 1996.

[15] P. Thompson and A. W. Toga, "A surface-based technique for warping three-dimensional images of the brain," *IEEE Transactions on Medical Imaging*, vol. 15(4), pp. 402–417, 1996.

[16] P. M. Thompson, D. Macdonald, M. S. Mega, C. J. Holmes, A. C. Evans, and A. W. Toga, "Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces," *Journal of Computer Assisted Tomography*, vol. 21(4), pp. 567–581, 1997.

[17] P. M. Thompson and A. W. Toga, "Detection, visualization and animation of abnormal anatomical structure with a deformable brain atlas based on random vector field transformations," *Medical Image Analysis*, vol. I(4), pp. 271–294, 1997.

[18] U. Grenander, M. I. Miller, and G. E. Christensen, "Deformable anatomical data bases using global shape models," in *1992 EIHB Workshop*, 1992.

[19] U. Grenander and M. I. Miller, ," in *Computational Anatomy: An Emerging Discipline*, 1998.

[20] S. Minoshima, R. A. Koeppe, K. A. Frey, and D. E. Kuhl, "Anatomic standardization - linear scaling and nonlinear warping of functional brain images," *Journal of Nuclear Medicine*, vol. 35(9), pp. 1528–1537, Sept. 1994.

[21] Y. Amit, "A non-linear variational problem for image matching," *Siam Journal of Scientific Computation*, vol. 15, no. 1, January 1994.

[22] J. Boes, B. Kim, K. A. Frey, and C. Meyer, "Mri-pet head registration via iterative minimization of mutual information," in *Proceedings of the 2nd Int.Conf. on Functional Mapping of the Human Brain*. June 1997, p. 149, Academic Press.

[23] B. Kim, J. Boes, K. A. Frey, and C. Meyer, "Mutual information for automated unwarping of rat brain autoradiographs," *Neuroimage*, vol. 5(1), pp. 31–40, Jan. 1997.

[24] C. Studholme, D. L. G. Hill, J. Wong anf M. N. Maisey, and D. J. Hawkes, "Registration measures for automated 3d alignment of pet and intensity distorted mr images," in *Proceedings in Image Fusion and Shape Variability Techniques*. July 1996, pp. 186–193, IEEE Computer Society Press ISBN 0-8186-7367-2.

[25] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Incorporating connected region labelling into automated image registration using mutual information," in *Proceedings of the IEEE workshop on Mathematical Methods in Biomedical Image Analysis, San Fransisco CA*. June 1996, pp. 23–31, IEEE Computer Society Press ISBN 0-8186-7367-2.

[26] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positiron emission tomography brain images by multiresolution optimisation of voxel similarity measures," *Medical Physics*, vol. 24, pp. 24–35, 1997.

[27] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the baysian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. Pattern Analysis and Machine Intelligence, Nov. 1984.

[28] S. Lakshmanan and H. Derin, "Gaussian markov random fields at multiple resolutions," *Markov Random Fields, Theory and Application*, pp. 131–157, 1993.

[29] G. E. Christensen, M. I. Miller, and M. Vannier, "A

deformable magnetic resonance textbook bason on elasticity," in *Application of Computer Vision in Medical Image Processing.* March 1994, Stanford University.

[30] G. E. Christensen, R. D. Rabitt, and M. M. Miller, "A deformable neuroanatomy based on viscous fluid mechanics," in *Proc. of the 1993 Conference on Information Sciences and Systems*, Mar. 1993, pp. 211–216.

[31] G. E. Christensen, R. D. Rabitt, and M. M. Miller, "3d brain mapping using a deformable neuroanatomy," *Physics in Medicine and Biology*, vol. 39, pp. 609–618, 1994.

[32] G. E. Christensen, R. D. Rabitt, and M. M. Miller, "Deformable templates using large deformation kinematics," *IEEE Transactions of Image Processing*, vol. 5(10), pp. 1435–1447, 1996.

[33] M. I. Miller, G. E. Christensen, Y. Amit, and U. Grenander, "Mathematical textbook of deformable neuroanatomies," in *Proc. of the National Academy of Science*, December 1993, vol. 90:24, pp. 11944–11948.

[34] C. Gramkow and M. Bro-Nielsen, "Comparison of three filters in the solution of the navier-stokes equation in registration," in *Proceedings of the Scandinavian Conference on Image Analysis (SCIA '97)*, 1997, pp. 795–802.

[35] K. Friston, J. Ashburner, C. D. Frith, J-B. Poline, J. D. Heather, and R. S. J. Frackowiak, "Spatial registration and normalization of images," *Human Brain Mapping*, vol. 2, pp. 165–189, 1995.

[36] D. L. Collins, G. Le Goualher, R. Venugopal, A. Caramanos, A. C. Evans, and C. Barillot, "Cortical constraints for non-linear cortical registration," *Lecture Notes in Computer Science, Visualization in Biomedical Comptuting*, pp. 307–316, 1996.

[37] I. Law, C. Svarer, and O. B. Paulson, "Characterization of the cortical responses during the performance of reflexive and antisaccadic eye movements," *Human Brain Mapping*, p. 1:323, 1995.

[38] T. R. DeGrado, T. G. Turkington, J. J. Williams, C. W. Stearns, J. M. Hoffman, and R. E. Coleman, "Performance characteristics of a whole-body pet scanner," *Journal of Nuclear Medicine*, vol. 35, pp. 1398–406, 1994.

[39] S. C. Strother, J. R. Anderson, K. A. Schaper, J. J. Sidtis, J. S. Liow, R. P. Woods, and D. A. Rottenberg, "Principal component analysis and the scaled subprofile model compared to intersubject averaging and statistical parametric mapping: I. "Functional connectivity" of the human motor system studied with [15-o]water pet," *Journal of Cerebral Blood Flow and Metabolism*, vol. 15, pp. 738–753, 1995.

[40] J. R Moeller and S. C. Strother, "A regional covariance approach to the analysis of functional patterns in positron emission tomographic data," *Journal of Cerebral Blod Flow Metabolism*, vol. 11, pp. A125–135, 1991.

[41] J. R Moeller, S. C. Strother, J. J. Sidtis, and D. A. Rottenberg, "Scaled subprofile model: A statistical approach to the analysis of functional activation patterns in positron emmision tomographic data," *Journal of Cerebral Blod Flow Metabolism*, vol. 7(5), pp. 649–658, 1987.

[42] K. V. Mardia and J. T. Kent, *Multivariate Analysis*, Academic Press, 1979.

[43] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T Vetterling, *Numerical Recipes*, Cambridge University Press, 1986.

[44] S. C. Strother, J. R. Anderson, K. A. Schaper, J. S. Sidtis, and D. A. Rottenberg, "Linear models of orthogonal subspaces and activated networks from functional activation pet studies of the human brain," in *Computational Imaging and Vision, Information Processing in Medical Imaging.* 1995, vol. 3, pp. 299–310, Kluwer Academic Pubs.

[45] S. C. Strother, J. R. Anderson, J. J. Sidtis, K. A. Schaper, L. K. Hansen, J. S. Liow, and D. A. Rottenberg, "Functional image registration metrics based on orthogonal subspaces," *Mapping and Modeling of the Human Brain*, 1997.

[46] B. Lautrup, L. K. Hansen, I. Law, N. Mørch, C. Svarer, and S. Strother, "Massive weight sharing: A cure for extremely ill-posed problems," in *Proceedings of Workshop on Supercomputing in Brain Research: From Tomography to Neural Networks: From tomography to neural networks, HLRZ, KFA Jülich, Germany*, H. J. Hermann, D. E. Wolf, and E. P. Pöppel, Eds. Nov. 1995, pp. 137–148, World Scientific.

[47] S. C. Strother, N. Lange, R. L. Savoy, J. R. Anderson, J. J. Sidtis, L. K. Hansen, P. A. Bandettini, K. O'Craven, M. Rezza, B. R. Rosen, and D. A. Rottenberg, "Multidimensional state spaces for fmri and pet activation studies," *Neuroimage*, vol. 3 (part 2), pp. S98, 1996.

[48] K. Friston, J. B. Poline, A. P. Holmes, C. D. Frith, and R. S. J. Frackowiak, "A multivariate analysis of pet activation studies," *Human Brain Mapping*, vol. 4, pp. 140–151, 1996.

[49] S. C. Strother, N. Lange, J. R. Anderson, K. A. Schaper, K. Rehm, L. K. Hansen, and D. A. Rottenberg, "Activation pattern reproducibility: Measuring the effects of group size and data analysis models," *Human Brain Mapping*, vol. 5, pp. 312–316, 1997.

[50] N. Mø rch, L. K. Hansen, S. C. Strother, C. Svarer, D. A. Rottenberg, B. Lautrup, R. Savoy, and O. B. Paulson, "Nonlinear versus linear models in functional neuroimaging: Learning curves and generalization crossover," in *Lecture Notes in Computer Science 1230: Information Processing in Medical Imaging*, J. Duncan and G. Gindi, Eds. 1997, pp. 259–270, Springer-Verlag.

M.Sc.E.E.Ph.D. Ulrik Kjems born March 1971, he obtained his Master of Science degree in Electrical Engineering at the Technical University of Denmark (DTU) February 1995 working with satellite orbit modelling, and his Ph.D. in signal processing at the Section for Digital Signal Processing, Department of Mathematical Modelling, also at DTU. His major research interrests have been working with methods for incorporating anatomical prior information into functional brain images and deformable models of the brain. In addition, his research interests are parameter estimation in Markov random fields and (non-)linear modelling of functional activation patterns in the human brain. See `http://imm.dtu.dk/~ uk` for more information and downloading of the warp software.