Groupwise Registration of Multimodal Images by an Efficient Joint Entropy Minimization Scheme

Žiga Špiclin, Boštjan Likar, and Franjo Pernuš

***Abstract*—Groupwise registration is concerned with bringing a group of images into the best spatial alignment. If images in the group are from different modalities, then the intensity correspondences across the images can be modeled by the joint density function (JDF) of the cooccurring image intensities. We propose a so-called *treecode registration method* for groupwise alignment of multimodal images that uses a hierarchical intensity-space subdivision scheme through which an efficient yet sufficiently accurate estimation of the (high-dimensional) JDF based on the Parzen kernel method is computed. To simultaneously align a group of images, a gradient-based joint entropy minimization was employed that also uses the same hierarchical intensity-space subdivision scheme. If the Hilbert kernel is used for the JDF estimation, then the treecode method requires no data-dependent bandwidth selection and is thus fully automatic. The treecode method was compared with the ensemble clustering (EC) method on four different publicly available multimodal image data sets and on a synthetic monomodal image data set. The obtained results indicate that the treecode method has similar and, for two data sets, even superior performances compared to the EC method in terms of registration error and success rate. The obtained good registration performances can be mostly attributed to the sufficiently accurate estimation of the JDF, which is computed through the hierarchical intensity-space subdivision scheme, that captures all the important features needed to detect the correct intensity correspondences across a multimodal group of images undergoing registration.**

***Index Terms*—Entropy, groupwise, image registration, joint density function (JDF), multimodality, multiscale, space partition.**

I. Introduction

REGISTRATION of images is needed in most of the applications that perform automated image analysis. Such applications are now well established in various fields ranging from medical imaging and remote sensing to computer vision. To perform a joint automated analysis of multiple images that were acquired by using different sensors, by varying illumination setups, at different times and/or viewpoints etc., these images must be spatially aligned. Thus, image registration should be a versatile tool that enables accurate, robust, and time-efficient spatial alignment of two or even more images.

Manuscript received December 19, 2010; revised May 15, 2011, October 20, 2011, and January 16, 2012; accepted January 16, 2012. Date of publication January 26, 2012; date of current version April 18, 2012. This work was supported in part by the Ministry of Higher Education, Science, and Technology of the Republic of Slovenia under Grant L2-2023, Grant L2-4072, and Grant J2-2246 and in part by the European Union under Grant ESRR-07-13-EU. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sina Farsiu.

The authors are with the Laboratory of Imaging Technologies, Department of Electrical Engineering, University of Ljubljana, 1000 Ljubljana, Slovenia (e-mail: ziga.spiclin@fe.uni-lj.si; bostjan.likar@fe.uni-lj.si; franjo.pernus@fe.uni-lj.si).

Digital Object Identifier 10.1109/TIP.2012.2186145

Scene information acquired from multiple sources or instances is generally represented by images, which are not directly comparable in intensities. In medical imaging for example, different insight about the underlying patient anatomy and/or function is offered by different acquisition techniques such as magnetic resonance (MR), computed tomography (CT), position emission tomography (PET), and ultrasound (US). In remote sensing, images capture the radiation intensity at different spectral bands of light, whereas in computer vision, changing the viewpoint or illumination direction diversely exposes the depth and shape information on the imaged scene. Such diverse scene information is likely to yield intensity uncorrelated images known as *multimodal* images. Bringing into spatial alignment two or more multimodal images presents a difficult image registration problem.

The intensity correspondences needed to guide the registration process can be established by assuming some form of a functional relationship between the intensity values of corresponding structures in the images. This functional relationship can either be explicit in a form of intensity transformation model [1] or implicit in a form of a JDF of image intensities [2]. The former seems less attractive since it needs to be parameterized specifically for the nature of the images being registered, which is contrary to the JDF that explains the statistical dependence between the image intensities. The JDF forms the basis of the state-of-the-art multimodal image registration methods [3], [4]. However, the dimensionality of the JDF grows by the number of images being registered. For the purpose of the groupwise image registration of multimodal images, an efficient but robust method is needed to estimate the high-dimensional JDF.

In this paper, we propose a so-called *treecode registration method* for registering a group of multimodal images that estimates the JDF through an efficient hierarchical subdivision of the joint intensity space (JIS). All images are simultaneously considered in the registration process, which is formulated as an iterative least-square gradient-based minimization of the joint entropy. The treecode method is compared with the ensemble clustering (EC) method [5] using similar image data sets and similar testing conditions. The main difference between the two methods is in the estimation of the JDF. In the EC method, a parametric Gaussian mixture model (GMM) is used, where the number of mixture components corresponding to the number of modes in the JDF has to be determined in advance either manually or automatically. This parameter critically defines the performances of the EC method [5]. Instead of modeling the clusters explicitly, the treecode method implements a Parzen kernel density through the hierarchical intensity-space subdivision scheme to obtain a representative high-dimensional JDF estimate.

II. Background

The main advantage of the emerging groupwise registration methods over the common pairwise registration methods [6] is that the correspondences between each pair of images depend on the joint information from the entire group of images. Depending on how this joint information is managed through the registration process, groupwise methods can be classified into sequential and simultaneous approaches.

Sequential approaches usually require a common reference or a *template* to which all images are registered in a pairwise manner. In [7], a mean geometry template was estimated through multidimensional scaling of unaligned images, whereas Marsland *et al.* [8] and Cootes *et al.* [9] applied the minimum description length principle in the search of a template that would most effectively explain the images and the pairwise transformations that are needed to align the images. Wang *et al.* [10] hierarchically cluster the images into a pyramid of classes and perform groupwise registration of similar images within each class at the bottom of the pyramid. Then, the center images of different classes are aligned from the bottom to the top of the pyramid. A similar strategy using a minimal spanning tree (MST) to find image pairs for registration was used in [11]. While effective for groupwise registration, these methods suffer from high computational complexity and have not been demonstrated to work with multimodal images.

More efficient groupwise registration methods are based on the simultaneous approach, in which a global criterion measures the information coherence across a group of images. The image group is then iteratively aligned by optimizing this global criterion with respect to (w.r.t) transformations of each image in the group. Among the first, Woods *et al.* [12] proposed added sums of squared intensity differences between all possible image pairs as the global criterion. Congealing framework [13] introduced the elementwise intensity entropy criterion that has been adopted for groupwise registration of MR images of the same acquisition protocol in [14]. They computed the so-called *stack entropy* from a density function of a stack of intensity values at each image location. To improve the correspondence detection, Wang *et al.* [15] used *attribute vectors* to represent contextual information about each image location in a form of local intensity density functions. This approach was further improved in [16] by allowing fuzzy correspondences in the spatial domain and across images in the group. If the intensity transformations between images are nonmonotone functions, as is usually the case for multimodal images, the mentioned global criteria might not measure the true intensity correspondences across the image group. Hence, these methods cannot be used to align multimodal images.

True multimodal groupwise image registration methods focus on the direct analysis and manipulation of the cooccurring intensities in the JIS, which spans the JDF. The rationale behind these methods is that the dominant cooccurring intensities form clusters in JIS. These clusters represent the most probable intensity correspondences between the images.

If the same clusters can be approximately identified when the images are misaligned, then this information can be used to guide the image registration process. Studholme and Cardenas [17] approximated the high-dimensional JDF, with the number of dimensions equal to the number of images being registered, by first finding the dominant clusters of cooccurring intensities and then replacing the clusters with weighted Gaussian functions. Their global criterion was a sum of total self-information (an asymptotic variant of the joint entropy) and two penalty terms enforcing geometric constraints on the registration solution. A similar technique based on high-dimensional histogramming was proposed in [4], in which the intensities were first binned and then the frequencies of only the nonempty histogram bins were computed. A multimodality extension to mutual information is proposed that uses the JDF estimate based on the high-dimensional histogram to align the images.

More recent methods have applied clustering or space partitioning techniques to capture the distribution of intensity points in JIS. Orchard and Jonchery [18] used normalized cuts to initialize the clusters in JIS, which represented a set of regressors, and then minimized generalized Euclidean distance from the set of regressors. An improved version of their method is done in [5], in which the authors used clustering based on a GMM. Then, the log-likelihood of the GMM was maximized that, in effect, reduced dispersion of the intensity points in JIS. Another class of methods is based on the graph representations of intensity points in JIS [19], [20], in which the dispersion of intensity points was quantified by computing the edge length of a MST. The mentioned methods try to directly minimize the dispersion of intensity points in JIS, which is intimately linked to the concepts of information-theoretic criteria for multimodal image registration [2], [3].

We propose a method for groupwise registration of multimodal images called the treecode registration method that minimizes *directly* the joint entropy so as to *indirectly* reduce the dispersion of intensity points in JIS. We note here that groupwise registration based on the minimization of joint entropy has already been proposed [5], [14], [17]; however, previous methods are either not adequate for registering multimodal images or may require data-dependent tuning of its parameters. There are two main contributions in this paper: 1) the hierarchical intensity-space subdivision scheme is the essential breakthrough that enables a representative and efficient approximation of the high-dimensional JDF and the joint entropy by the Parzen kernel method; and 2) the use of the Hilbert kernel that captures all important features of the JDF without the need to tune the kernel bandwidth parameters, as with the common Gaussian kernel, for each specific image group undergoing registration.

III. Method

The proposed treecode method for groupwise registration of multimodal images is based on minimizing the joint entropy of the images by 1) using the Parzen kernel method to estimate the JDF, 2) evaluating the joint entropy, and 3) moving the images so as to minimize the joint entropy. The process of JDF estimation, joint entropy calculation, and motion adjustment is simultaneously executed through 4) an efficient hierarchical subdivision of the JIS.

The method takes as input an image group Z that consists of D images $Z_1, \ldots, Z_D$. Pixel coordinates in the overlapping domain $x \in \Omega_Z$ of images in Z define the D-dimensional vectors

of image intensities, which we refer to as *intensity points* and denote as $z_i \in \mathbb{R}^D$ for corresponding pixel coordinates $x(i)$, $i = 1, \ldots, N$. Thus, the kth row of intensity point z_i holds intensities of the kth image, which are referred to as $[z_i]_k$. Intensity points populate the JIS and are treated as independent and identically distributed random variables. A sample-based estimate of their JDF is required to compute the joint entropy.

The joint entropy is closely related to mutual information [2], [3], originally defined between two images as

$$\mathrm{MI}(Z_1, Z_2) = H(Z_1) + H(Z_2) - H(Z_1, Z_2) \tag{1}$$

where $H(Z_k) = -\int p([z]_k) \log p([z]_k) dz_k$ is the univariate entropy values and $H(Z_1, Z_2) = -\int p(z) \log p(z) dz$ is the joint entropy. $p([z]_k)$ and $p(z)$ are the univariate and the bivariate density functions, respectively. A straightforward extension of mutual information for more than two variables is known as the total correlation [21], which in the case of D images reads

$$\mathrm{MI}(Z) = \sum_{k=1}^{D} H(Z_k) - H(Z). \tag{2}$$

The total correlation measures the degree of statistical dependence of a group of variables. In the image registration case, the highest statistical dependence is expected when all images are best aligned. A strategy for groupwise registration is thus to maximize the total correlation. In (2), the statistical dependence between images is directly expressed by the multivariate joint entropy, which acts as a lower bound of the total correlation. Instead of maximizing the total correlation, it is more simple to maximize its lower bound, i.e., to minimize the joint entropy of the images.

The next subsections describe the estimators of density, joint entropy, motion parameters, the hierarchical JIS subdivision method, and the implementation details.

A. Density Estimation

We will use the Parzen kernel method to estimate the true probability density $p(z)$ from the intensity points $z \in Z$. By definition of the Parzen kernel method [22], the true probability density $p(z)$ can be estimated as a superposition of kernel functions centered on M samples z_j drawn from Z

$$p(z) \approx p^*(z) = \frac{1}{M} \sum_{j=1}^{M} K_h(z - z_j) \tag{3}$$

where the kernel function $K_h(z)$ integrates to 1, $K_h(z) = 1/(h^D) K(z/h)$, and h is the smoothing parameter known as the bandwidth [23]. The most common kernel function $K_h(z)$ is the (multivariate) Gaussian

$$K_{\mathcal{G}|\psi}(z) = \sqrt{(2\pi)^{-D} |\psi|^{-1}} \exp\left(-\frac{1}{2} z^T \psi^{-1} z\right) \tag{4}$$

where ψ is the covariance matrix, typically defined as $\psi = \mathrm{diag}([h_1^2, \ldots, h_D^2]^T)$. We will also use the Hilbert kernel proposed in [24], which is defined as

$$K_{\mathcal{H}}(z) = [V_D \log M]^{-1} \sqrt{(z^T z)^{-D}} \tag{5}$$

where V_D is the volume of the unit ball in $\mathbb{R}^D$. Note that h is canceled in $K_h(z/h)$, where $K_h(z)$ is defined as $K_h(z) = 1/(h^D)\, 1/\|z\|^D$, hence, the Hilbert kernel has no bandwidth parameters. The density estimate based on the Hilbert kernel is weakly consistent at almost all z, i.e., $p^*(z) \to p(z)$ in probability at almost all z [24].

B. Joint Entropy

By using the Parzen kernel density estimate $p^*(z)$ the joint entropy can be computed as $H(Z) = -\mathbb{E}_z[\log p^*(z)]$. This expression is difficult to evaluate; therefore, we will approximate the joint entropy by a sample mean $H(Z) \approx -1/N \sum_{i=1}^{N} \log p^*(z_i)$ that converges to the true expectation at a rate proportional to $\sqrt{1/N}$ [3]. By using kernel functions in (4) and (5), the density estimate can be expressed as $p^*(z) = 1/V_K \sum_{j=1}^{M} K(z - z_j)$, where V_K is the kernel-dependent density normalization. In this way, the joint entropy is computed as

$$H^*(Z) = -\frac{1}{N} \sum_i \log \left[\frac{1}{M V_K} \sum_{j \neq i} K(z_i - z_j) \right] \tag{6}$$

where the case $j = i$ was excluded to avoid the singularity of the Hilbert kernel for $K_{\mathcal{H}}(0)$; hence, $M = N - 1$. This issue will be further addressed in Sections III-E and V-F.

In (6) and in the following equations, all sums run over all intensity points, except if otherwise indicated.

C. Motion Adjustment

Let θ represent the motion parameters of all images in the image group. During the image registration process, the intensity points z change w.r.t. the motion parameters θ, which we will denote as z^θ. We want to find such motion parameters θ that minimize the joint entropy $H^*(Z^\theta)$ and thus bring all D images into the best spatial alignment. Following [5], we propose an iterative solution using a Newton-type minimization step.

The minimal joint entropy $H^*(Z^\theta)$ w.r.t. motion parameters can be obtained by setting its gradient vector to zero as follows:

$$\frac{\partial}{\partial \theta} H^*(Z^\theta) = \mathbf{0}. \tag{7}$$

Registration of D images, each having P motion parameters, yields PD motion parameters in θ, i.e., the gradient vector in (7) is a $PD \times 1$ vector.

We set $d_{ij}^\theta = z_i^\theta - z_j^\theta$ and use (6) to rewrite the gradient vector of joint entropy as

$$\frac{\partial}{\partial \theta} H^*(Z^\theta) = \frac{1}{N} \sum_i \frac{\partial}{\partial \theta} \left[\log M V_K - \log \sum_{j \neq i} K\left(d_{ij}^\theta\right) \right]. \tag{8}$$

By partial differentiation, the right-hand side becomes

$$\frac{1}{N} \sum_i \sum_{j \neq i} W_{ij}^\theta R_K \frac{\partial d_{ij}^\theta}{\partial \theta} d_{ij}^\theta \tag{9}$$

where

$$W_{ij}^{\theta} = \frac{K\left(d_{ij}^{\theta}\right)}{\sum_{k\neq i} K\left(d_{ik}^{\theta}\right)} \tag{10}$$

is a scalar weight between 0 and 1. R_K is a kernel-dependent factor that evaluates to ψ^{-1} and $D\|d_{ij}^{\theta}\|^{-2}$ for the Gaussian and the Hilbert kernel, respectively. Lastly, $\partial d_{ij}^{\theta}/\partial\theta$ is a $PD \times D$ matrix of derivatives of D intensity values w.r.t. the PD motion parameters.

Improved motion parameters that set expression (9) to a zero vector can be found by using a linear approximation of the spatial transformation

$$d_{ij}^{\theta+\Delta\theta} = d_{ij}^{\theta} + \frac{\partial d_{ij}^{\theta}}{\partial\theta}^{T} \Delta\theta \tag{11}$$

where $\Delta\theta$ is a small step in the motion parameters. Note that, in d_{ij}^{θ}, both z_i^{θ} and z_j^{θ} are dependent on the motion parameters. We replace d_{ij}^{θ} in (9) by a linear approximation in (11) and obtain a linear equation in $\Delta\theta$

$$\sum_i \sum_{j\neq i} W_{ij}^{\theta} R_K \frac{\partial d_{ij}^{\theta}}{\partial\theta}\left[d_{ij}^{\theta} + \frac{\partial d_{ij}^{\theta}}{\partial\theta}^{T}\Delta\theta\right] = \mathbf{0}. \tag{12}$$

By factoring out $\Delta\theta$, we get a linear system, i.e.,

$$\left[\sum_i \sum_{j\neq i} W_{ij}^{\theta} R_K \frac{\partial d_{ij}^{\theta}}{\partial\theta}\frac{\partial d_{ij}^{\theta}}{\partial\theta}^{T}\right]\Delta\theta = \left[\sum_i \sum_{j\neq i} W_{ij}^{\theta} R_K \frac{\partial d_{ij}^{\theta}}{\partial\theta} d_{ji}^{\theta}\right] \tag{13}$$

or

$$A\Delta\theta = b \tag{14}$$

where A is a $PD \times PD$ system matrix, and b is the $PD \times 1$ response vector, as defined in (13). This linear system can be easily solved to obtain the motion step $\Delta\theta$, which is used to update the current motion parameters θ.

D. Hierarchical Joint Intensity-Space Subdivision

The computational complexity of (13) is $\mathcal{O}(N^2P^2D^3)$, typically with $N \gg \max(P, D)$. To break the $\mathcal{O}(N^2)$ complexity due to the double sum in (13), we propose an efficient $\mathcal{O}(N \log N)$ scheme based on a tree-structured recursive hierarchical subdivision of the JIS into hypercubes [25]. The idea is to group together increasingly large groups of intensity points at increasingly large intensity differences. In this way, interactions between distant groups of intensity points in JIS can be efficiently approximated without summing over each pair of intensity points. Overall, the computational complexity can thus be reduced to $\mathcal{O}(N \log NP^2D^3)$.

We start the space subdivision with an empty hypercubic cell that is big enough to contain all intensity points. Then, intensity points are sequentially loaded into this *root* cell. If more than one intensity point is contained in the same cell, that cell is divided into 2^D equally sized hypercubic subcells. This process of subdividing is continued to as high a level as required. These cells are then grouped together into larger hypercubic cells up to the root cell, forming a hierarchical tree structure. The final step in constructing the tree is to tag each cell B with

$$N_B = \sum_{k\in B} N_k$$

$$\widehat{z}_B^{\theta} = \frac{1}{N_B}\sum_{k\in B} N_k\, z_k^{\theta} \tag{15}$$

$$\widehat{E}_B^{\theta} = \frac{1}{N_B}\sum_{k\in B} N_k \frac{\partial z_k^{\theta}}{\partial\theta} \tag{16}$$

$$\widehat{F}_B^{\theta} = \frac{1}{N_B}\sum_{k\in B} N_k \frac{\partial z_k^{\theta}}{\partial\theta}\, z_k^{\theta} \tag{17}$$

$$\widehat{G}_B^{\theta} = \frac{1}{N_B}\sum_{k\in B} N_k \frac{\partial z_k^{\theta}}{\partial\theta}\frac{\partial z_k^{\theta}}{\partial\theta}^{T} \tag{18}$$

where N_k is the number of intensity points in each cell or subcell of B. By propagating tags from the bottom-level cells up to the tree root, a tagged tree may be obtained in time of $\mathcal{O}(N \log NP^2D^3)$.

For each intensity point, we evaluate the sum over j's in the bracketed terms of (13) by a simple recursive calculation starting at the root cell of the tree. Let l_B be the length of cell B being processed and d_{iB}^{θ} the difference between the cell's center-of-mass $\widehat{z}_B^{\theta}$ to intensity point z_i^{θ}. If $l_B/d_{iB}^{\theta} \leq \phi$, where ϕ is a fixed accuracy threshold, then we compute

$$A_{iB} = W_{iB}^{\theta} R_K \left[G_i^{\theta} + \widehat{G}_B^{\theta} - \widehat{E}_B^{\theta}\, E_i^{\theta T} - E_i^{\theta}\widehat{E}_B^{\theta T}\right] \tag{19}$$

$$b_{iB} = W_{iB}^{\theta}\, R_K \left[\widehat{E}_B^{\theta}\, z_i^{\theta} + E_i^{\theta}\widehat{z}_B^{\theta} - F_i^{\theta} - \widehat{F}_B^{\theta}\right] \tag{20}$$

where A_{iB} is the partial system matrix and b_{iB} is the partial response vector. Note that R_K is computed as $\|d_{iB}^{\theta}\|^{-2}$ for the Hilbert kernel. While traversing the tree structure, A_{iB} and b_{iB} are added to the total system matrix A and the total response vector b in (14), respectively. E_i^{θ}, F_i^{θ}, and G_i^{θ} represent the respective cell tags, as in (16)–(18), for the case of a cell containing a single intensity point.

For $l_B/d_{iB}^{\theta} > \phi$, the current cell B is resolved into its 2^D subcells, which are then recursively examined one by one. This process continues down to the bottom-level tree cells and is of $\mathcal{O}(\log NP^2D^3)$. The JIS subdivision and the interpretation of the cell acceptance criterion can be observed in Fig. 1.

E. Implementation

The proposed method was implemented in Matlab (Mathworks Inc., Natick, MA). The pseudocode in Algorithm 1 summarizes the main computational steps of the proposed method. Image groups in our registration experiments were successively registered at scales 1:16, 1:8, 1:4, and 1:2 of the original resolution and at the original image resolution. Motion parameters obtained at one scale were used as initial motion parameters on the successive higher scale.

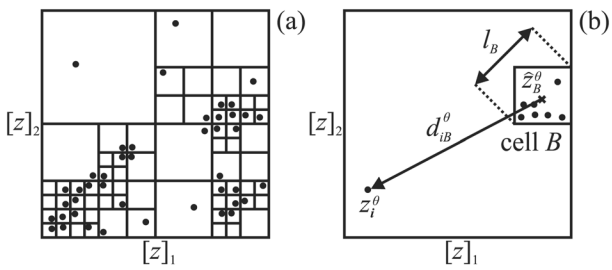


Fig. 1. (a) Two-dimensional JIS with a corresponding cell subdivision and (b) cell interaction approximation between cell B with its center-of-mass $\hat{z}_B^\theta$ and intensity point z_i^θ.

For resolution scales that yielded a large number of intensity points, we randomly sampled the intensity points in each iteration of the registration process so as to limit the maximal number of intensity points to 2000; hence, $N^* = \min(2000, N)$. This random sampling scheme introduces some noise in the gradient estimation in (9); however, as shown in the Section V, the proposed method still registers images close to their gold-standard position. Moreover, similar random sampling schemes were used in [3] and [14] to improve the computational efficiency of the registration method and to effectively escape from the local minima in the registration criterion.

The covariance matrix ψ of the Gaussian kernel density estimate in (4) was set according to the maximal smoothing principle [23] as follows:

$$\psi = \left[\frac{2^{-D}\,(D+8)^{(D+6)/2}}{16\,N\,\Gamma\left(\frac{D+8}{2}\right)\,(D+2)}\right]^{2/(D+4)} \Sigma^* \qquad (21)$$

where Σ^* is the diagonal N^*-sample covariance, i.e., $\Sigma^* = \mathrm{diag}([\sigma_1^2, \ldots, \sigma_D^2]^T)$, and $\Gamma(\cdot)$ denotes the Gamma function. Contrary, the Hilbert kernel in (5) has no bandwidth parameters. However, to guarantee the numerical stability of $K_{\mathcal{H}}(z)$ for $z \to 0$, we regularize the L_2-norm in (5) as follows:

$$K_{\mathcal{H}|\varepsilon}(z) = [V_D\ \log M]^{-1}\sqrt{(z^T z + \varepsilon^2)^{-D}}. \qquad (22)$$

The regularization parameter ε must be set so as to renormalize via (10) the influence of the intensity points closest in distance. Since, for $\varepsilon = 0$, the individual intensity points closest in distance dominate the motion step direction computed from (19) and (20), ε can be set w.r.t. the average minimal distance between the intensity points, i.e., $\varepsilon \sim \mathbb{E}[\min_j(d_{ij}^\theta)]$; $i, j = 1, \ldots, N$, $i \neq j$. The robust estimator of this distance can be derived from the distribution of hypercubic cells B w.r.t. the tree level L, i.e., $\#B(L)$, obtained by the hierachical space subdivision scheme proposed in Section III-D. We take the tree level at the 95th percentile of $\#B(L)$, denoted as $L_{0.95}$, and set ε according to the corresponding cell diagonal as

$$\varepsilon = l_0 D^{1/2} 2^{-L_{0.95}} \qquad (23)$$

where l_0 is the sidelength of the root cell. The hierarchical JIS subdivision had a maximum of 32 levels, whereas the cell acceptance threshold ϕ was set to 1. The influence of selecting parameters ε, ϕ, and N^* on the performances of the treecode method is further examined in Section V-F.

Exceptions to the given testing conditions and details about experiments are given in Section IV.

Algorithm 1: Pseudocode of the treecode registration method

input: initial image group Z_0
input: initial motion parameters θ
for each scale **do**
  $Z_{\mathrm{scaled}} \leftarrow$ scale image group Z_0
  $Z \leftarrow$ apply motion (θ) to image group Z_{scaled}
  repeat
    $Z^* \leftarrow$ sample N^* feature points from Z
    $\mathcal{T} \leftarrow$ create tree using Z^* (see Section III-D)
    for each feature point Z^* **do**
      $\mathcal{B} \leftarrow$ init cell list and insert *root* cell of $\mathcal{T}$
      repeat
        $B \leftarrow$ retrieve new cell from $\mathcal{B}$
        if $l_B / d_{iB}^\theta \leq \phi$ **then**
          compute A_{iB} and b_{iB}
          $A \leftarrow A + A_{iB}$
          $b \leftarrow b + b_{iB}$
        else
          $\mathcal{B} \leftarrow$ insert subcells $\mathcal{T}(B)$
        end if
      until cell list $\mathcal{B}$ empty
    end for
    $\Delta\theta \leftarrow$ motion adjustment (see Section III-C)
    $\theta \leftarrow \theta + \Delta\theta$
    $Z \leftarrow$ apply motion (θ) to image group Z_{scaled}
  until converged ($\Delta\theta$ is small)
end for
output: Z is registered image group at full scale
output: θ holds optimal motion parameters

IV. Experiments

The proposed treecode method was tested and compared with the EC method [5] so as to demonstrate that: 1) the treecode method has similar and, for some data sets, even superior performances as the EC method; and 2) the proposed method requires no data-dependent tuning of its parameters. To enable a direct comparison between the two methods, the outline of the experiments and results sections closely follows the one in [5]. The EC method was run using the parameter values (i.e., the number of

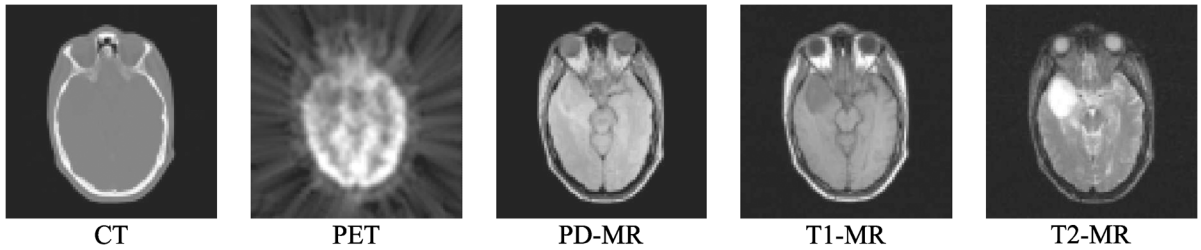


Fig. 2. Two-dimensional slice from the 3-D RIRE image group that was used to test the 3-D rigid-body registration.

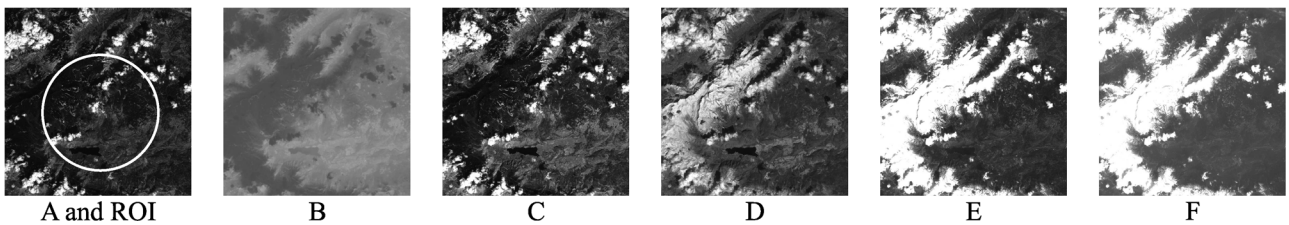


Fig. 3. Satellite images of different spectral bands of light that were used to test the 2-D affine registration. The ROI that was used for testing the image registration methods is outlined in image A.

Gaussians in GMM), as proposed in [5], since the experiments were carried out on similar image data sets and under similar testing conditions. The GMM was initialized by the k-means clustering algorithm. All experiments were executed using all images simultaneously in the registration process.

The criteria for evaluating and comparing the performances of the registration methods were also similar to the ones proposed in [5]. The estimated transformations, as output by the registration methods, were compared with the gold-standard transformations. *Registration error* was defined as the average pixel displacement from its true (gold standard) position, which is computed for each image pair in the data set and over all pixels used in the registration. The accuracy of the registration methods was measured by the *average registration error*, which is computed only for the successful registrations. Registration of an image pair was considered successful if the registration error was below three pixels. The *image-pair success rate* was defined as the overall percentage of successfully registered image pairs, whereas the *trial success rate* was defined as the percentage of trials that successfully registered all image pairs. To compare the accuracy of the methods, the P values of the one-sided paired t-test were computed between the corresponding means of the average registration errors of the successful trials. Let A and B be the methods under comparison for which $\mu_A > \mu_B$; then, the null hypothesis was defined as $H_0 : \mu_A = \mu_B$ and the alternative hypothesis as $H_1 : \mu_A > \mu_B$. The significance level was set at $\alpha = 0.05$. The reported P values, which are denoted as $P_{A=B}$, refer to the likelihood of the null hypothesis.

Four publicly available multimodal image data sets and a synthetic monomodal image data set were used to test the performances of the registration methods. Each image data set was first registered by a supplied gold-standard registration. Then, known displacements were applied to the registered images to obtain multiple trial image data sets to run the registration experiments. Details about image data sets and the registration experiments are given in the following subsections.

A. 3-D Medical Volumes

Medical volumes obtained from the Retrospective Image Registration Evaluation (RIRE) project's training data set [26] were used to test the 3-D rigid-body registration. Fig. 2 shows a slice from the aligned image group. The group consists of a CT, PET, and three MR volumes (PD-, T1- and T2-weighted). The volumes were initially aligned by ground-truth registration supplied by the RIRE project, then rescaled and zero padded to $80 \times 80 \times 32$ isotropic voxels. Refer to [5] for details. One hundred trial image groups were generated by applying randomly three translations and three rotations to each image. The random translations and rotations were chosen from the range $[-5, 5]$ pixels and degrees, respectively. A multiresolution registration was carried out at scales of 1:4 and 1:2 and at the original resolution.

B. Satellite Images

Landsat 7 satellite images were used to test the affine registration. Images were obtained from the Landsat website (http://landsat.gsfc.nasa.gov) using as parameters the latitude 46.0 and longitude 13.6 while the acquisition date was 29 April 2001. A 800×800 pixels ROI was cut out of the original Landsat 7 images, with the upper-left corner at (4600, 2200). The Landsat 7 spectral-band images 7, 6H, 5, 4, 3, and 1 are shown in Fig. 3 as images A to F, respectively. One hundred trial image groups were generated by applying to the images a random six degree-of-freedom (DOF) affine displacements. Parameter values for scale [0.95, 1.05], shear $[-0.2, 0.2]$, translation $[-5, 5]$, and rotation $[-5\ 5]$ were uniformly sampled from the given ranges. Note that the translations and rotations are given in pixels and degrees, respectively. The overlap region used for registration is shown in the image A in Fig. 3.

C. Variable Illumination

Face images illuminated from different angles (far left to far right) are shown in Fig. 4 and were obtained from the Extended Yale Face Database B [27]. One hundred trial image

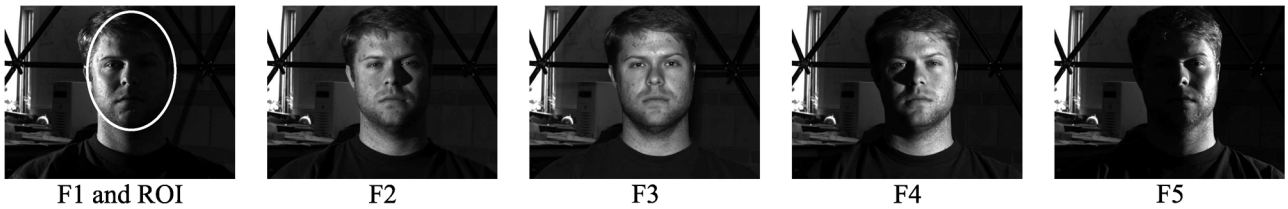


Fig. 4. Images from the Extended Yale B Face Database, in which faces were illuminated from different angles. The ROI that was used for testing the image registration methods is outlined in image F1.

groups were assembled by generating random rigid-body displacements of the images by $[-10, 10]$ pixels or degrees. The overlap region marked by ROI in the image F1 of Fig. 4 was used for registering the images.

D. Monomodal Time Series

A figure of merit for the groupwise methods is also the number of images that can be simultaneously registered. Therefore, a time series of ten simulated functional MRI (fMRI) images was composed by adding artificial activation to the occipital and temporal regions of a single 80×80 pixels T2-weighted MR the brain image. The artificial activations were generated using the fMRI Simulation Toolbox (SimTB) (http://mialab.mrn.org). The parameters of the fMRI simulator were set to trigger a series of three events overlapping in the time domain over a period of 20 s, and the acquisition period was set to 2 s. One hundred trial image groups were generated by applying to each image a random translation and rotation from the range of $[-10, 10]$ pixels or degrees, respectively. Finally, all the images were corrupted by additive Gaussian noise that presented 5% of the average brain intensity.

Three methods were tested on the fMRI time series, namely the EC method [5], the congealing method (CG) [14], and the proposed method. The EC method was initialized with one Gaussian for the GMM, whereas in the CG, the bandwidth σ of the Gaussian kernels was set to 1, and the same random sampling scheme was used as for the proposed method. All the methods were run successively at scales of 1:4 and 1:2 and at the original resolution.

E. Nonrigid Registration

In the medical imaging community, images usually need to be aligned with a high-DOF spatial transformation so as to facilitate automatic image analysis. Image registration with a high-DOF spatial transformation is also known as nonrigid image registration [6]. To test the proposed treecode method in the nonrigid setting, we used three simulated MR images (T1-, T2, and PD-weighted) from the BrainWeb project [28] that contained artificial multiple sclerosis lesions, had a noise level of 3% relative to the brightest tissue, and a 20% intensity nonuniformity. From each of the $181 \times 217 \times 181$ volumes, only the axial 2-D slices #105 were used for testing the nonrigid groupwise registration (see Fig. 5).

Nonrigid groupwise image registration was based on thin-plate splines (TPSs) [29], in which the spatial transformation is defined by the displacement of *control points* from their reference position. We used a 3×3 grid of the TPS control points,

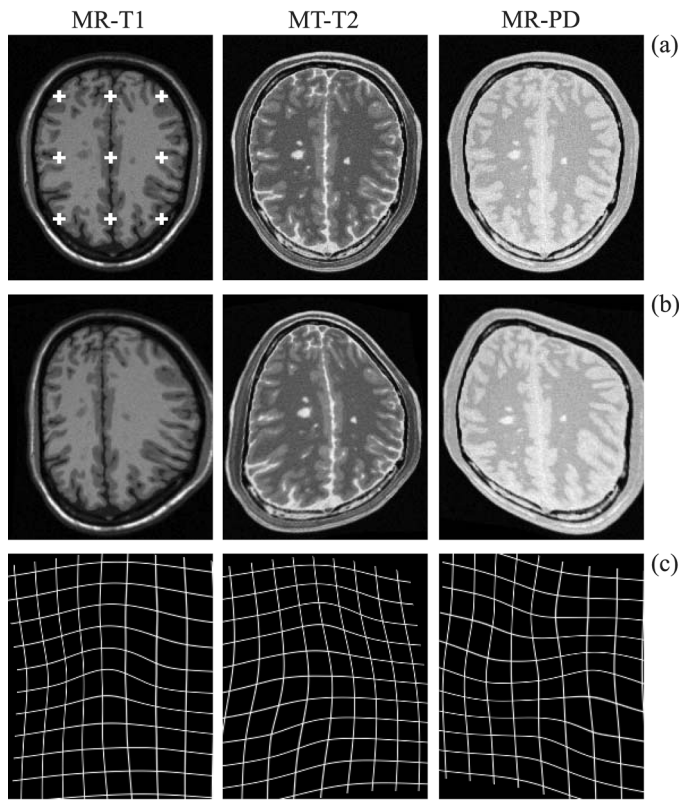


Fig. 5. (a) BrainWeb simulated MR images were used to test the nonrigid registration based on TPSs, using (top left) 3×3 control point grid. Example of (b) deformed test images and (c) corresponding deformation maps.

as depicted in Fig. 5 on T1 image, yielding 18 free parameters per image. One hundred trial image groups were generated by displacing control points' coordinates randomly in the range of $[-10, 10]$ pixels. The EC [30] and the proposed treecode methods were run successively at scales of 1:8, 1:4, and 1:2 and at the original resolution.

F. Properties and Performances of the Treecode Method

First, we assessed the accuracy of the approximate motion update $\Delta\theta^*$, as computed by the treecode method using (19) and (20), w.r.t. the number of intensity samples N and w.r.t. the cell acceptance threshold ϕ, which governs how point-to-group interactions are computed within the hierarchical intensity-space subdivision. For this purpose, trial image groups from the Yale Face Database were used (c.f. Section IV-C). Relative error between the approximate motion update $\Delta\theta^*$ and the true motion

TABLE I
SUMMARY OF IMAGE REGISTRATION RESULTS ON FIVE DIFFERENT DATA SETS BY RUNNING THE EC, THE CG, AND THE PROPOSED TREECODE REGISTRATION METHODS. THE TREECODE METHOD WAS TESTED WITH THE GAUSSIAN (TG) AND THE HILBERT (TH) KERNEL. NUMBERS IN BOLD INDICATE BEST RESULTS FOR EACH CRITERION AND DATA SET. STARS * INDICATE STATISTICALLY SIGNIFICANT DIFFERENCE IN ERROR MEANS FOR $\alpha = 0.05$ BETWEEN THE TREECODE AND OTHER METHODS. SEE FIG. 6 FOR P VALUES

	Mean average registration error [pixels]				Image-pair success rate [%]				Trial success rate [%]			
Dataset \ Method	EC	TG	TH	CG	EC	TG	TH	CG	EC	TG	TH	CG
RIRE (Sec. IV-A)	1.18	0.181	**0.159***		45.6	**100**	**100**		31	**100**	**100**	
Satellite (Sec. IV-B)	**0.055**	0.094	0.074		97.7	**100**	**100**		80	**100**	**100**	
Faces (Sec. IV-C)	**0.931***	1.03	1.03		**100**	99.6	99.6		**100**	99	99	
fMRI (Sec. IV-D)	**0.061***	0.089	0.087	0.083	**100**	98.9	98.9	**100**	**100**	97	97	**100**
BrainWeb (Sec. IV-E)	0.595	0.621	**0.119***		91.3	85.7	**100**		88	79	**100**	

update $\Delta\theta$, obtained by direct implementation of (13), was computed by

$$e_{\Delta\theta}[\%] = 100 \cdot \frac{\|\Delta\theta^* - \Delta\theta\|}{\|\Delta\theta\|}. \tag{24}$$

Moreover, the computational efficiency was evaluated by measuring the corresponding execution times.

Next, we studied the accuracy of the JDF as estimated by the treecode method. For this purpose, artificial JDFs were generated with a predefined number of modes R and dimensionality D as

$$p(z) = \frac{1}{R}\sum_{k=1}^{R} K_{\mathcal{G}|\psi_k}(z - \mu_k) \tag{25}$$

where mode positions μ_k and the covariances ψ_k were randomly chosen. Note that (25) is a randomly generated GMM. From each such JDF, N samples were drawn and used to reestimate the JDF by the treecode method and the expectation–maximization GMM (EM-GMM) with a fixed number of components R^*. Note that EM-GMM is used in the EC method. Pointwise accuracy of the estimated JDF was measured by the *integrated squared error* (ISE) [23]

$$\mathrm{ISE} = \int \left[p^*(z) - p(z)\right]^2 dz \tag{26}$$

while the overall accuracy was assessed by the *mean ISE* (MISE) as $MISE = \mathbb{E}[ISE(p^*, p)]$. MISE was evaluated for ten trial JDFs and assessed for various ranges of R, D, and N.

V. RESULTS

A. 3-D Medical Volumes

Results for the 3-D rigid-body registration of the RIRE multimodality medical volumes are shown in Table I and Fig. 6(a). The average registration error for the unregistered volumes was 6.6 pixels. In the successful trials, the proposed treecode method had an average registration error of 0.181 and 0.159 pixels for the Gaussian kernel and the Hilbert kernel, respectively, which was considerably lower than the average registration error of 1.18 pixels for the EC method. Treecode methods successfully registered all trials and all image pairs, whereas the respective percentages were much lower for the EC at 31% and 45.6%. The average times to register the volumes were 384 and 268 s for the EC and the treecode methods, respectively.

The trial success rate of the EC method was strongly influenced by the initial JDF estimate at 1:4 scale of resolution (see Fig. 6), where the number of intensity points for the JDF estimate was the lowest. If the initial JDF estimate could not locate the true intensity correspondences, then usually, all the images in the group failed to converge to their gold-standard position. In the EC method, one has to specify in advance the number of components for the GMM, which might not reflect the correct number of modes in the JDF. Moreover, the number of modes might change during the registration process. The proposed method measures the joint entropy directly from the intensity points and, thus, does not need to model the modes explicitly, which might explain a considerably higher trial success rate of the proposed method versus the EC method.

B. Satellite Images

One hundred trial image groups had a 19.9-pixel initial average registration error. The treecode method had an average registration error of 0.094 and 0.074 for the Gaussian kernel and the Hilbert kernel and has successfully registered all trials and all image pairs. The EC method had a slightly lower average registration error of 0.055 pixels, 80% successully registered trials, and 97.7% successfully registered image pairs [c.f. Table I and Fig. 6(b)]. The difference in the average registration errors between the methods is, however, statistically insignificant with $P_{\mathrm{TG=EC}} = 2.3 \cdot 10^{-1}$ and $P_{\mathrm{TH=EC}} = 1.7 \cdot 10^{-1}$, according to the paired t-test. The trial success rate was lower for the EC method mostly due to the image pair B–D (see Fig. 3) that failed to align in 17 trials (see Fig. 6). By adding only one Gaussian to the GMM (total of seven) in the EC method, the average registration error rose to 0.52 pixels, and the trial and image-pair success rates were 69% and 86.6%.

Average times to execute the registrations were 140 and 183 s for the EC and the treecode method, respectively. The treecode method spent most of the time (>90%) registering images in the last resolution scale. The motion parameters oscillated about the gold-standard position due to the random sampling scheme, and the registration was stopped when reaching the maximal number of iterations.

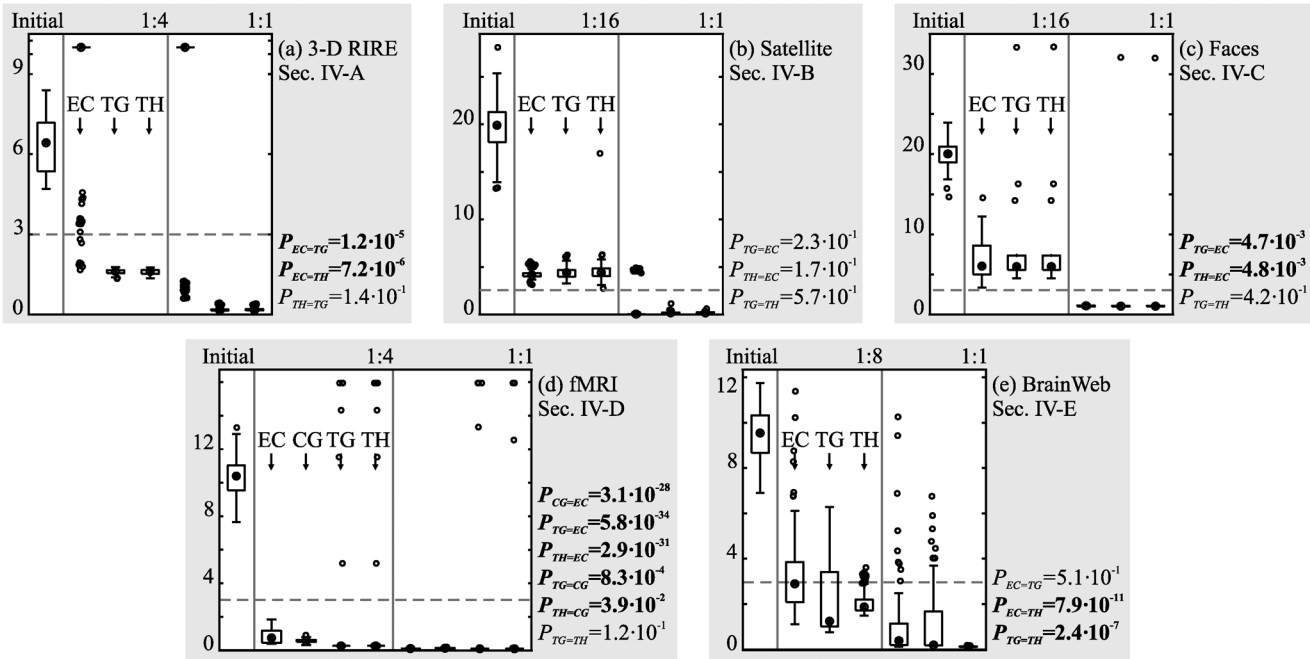


Fig. 6. Box–whiskers diagrams of the average registration errors for (a) 3-D RIRE, (b) Satellite, (c) Faces, (d) fMRI, and (e) BrainWeb image data sets. The initial average registration errors are plotted on the left of each box–whisker diagram, then the corresponding errors, as output by the registration methods at the initial and original resolution scales, are plotted to the right. The open dots represent those trials in which the average registration error was more than 1.5 of the interquartile range away from the median value (closed dot). The tested methods are referred to as EC, treecode with Gaussian (TG) and with Hilbert (TH) kernel, and CG. Corresponding P values for the one-sided paired t-tests are printed to the right of each box–whisker diagram (c.f. Section IV for definitions). Significance level is set at $\alpha = 0.05$; thus, for clarity, the P values <0.05 are marked in bold.

C. Variable Illumination

Results for the variable illumination experiment using face images are shown in Table I and Fig. 6(c). The initial average registration error for the unregistered images was 20.4 pixels. The EC method had an average registration error of 0.931 pixels and successfully registered all 100 trial image groups and all image pairs. If one Gaussian was added to the GMM, the average registration error was slightly worse at 0.96 pixels, whereas the trial and image-pair success rates dropped significantly to 76% and 82.4%. This indicates that the number of Gaussian components in the GMM has to be perfectly tuned to the registration problem at hand.

The treecode method had a higher average registration error at 1.03 pixel for both the Gaussian kernel and the Hilbert kernel. Similarly, using either of the kernels, 99% trial and 99.6% image-pair success rates were achieved. Average times to execute the registrations were 129 and 55 s for the EC and the treecode method, respectively.

D. Monomodal Time Series

Results for the fMRI time-series registration are shown in Table I and Fig. 6(d). The average registration errors for the successful registrations were 0.061 and 0.083 pixels for the EC and the congealing registration method. The treecode method had 0.089 and 0.087 pixels of average registration error for the Gaussian kernel and the Hilbert kernel, respectively. Both the EC and the CG successfully registered all trials and image pairs, whereas the proposed method successfully registered 97% of trials and 98.9% of image pairs, regardless of the kernel function used. The average times for a trial image group to register were 1.81, 32, and 145 s for the EC, the treecode method, and the CG, respectively.

Surprisingly, small average registration errors and fast registration time of the EC method result from using only one Gaussian for the GMM. Using only one Gaussian tricks the GMM to model only the intensity points in the brain overlap region that form a distinct cluster near the diagonal of JIS. Thus, other intensity points that may distract the registration are effectively ignored. We found that, by using two or three GMM components, the accuracy drops to 0.17 or 0.39, respectively. To set the optimal number of Gaussians for the GMM thus requires *a priori* analysis of the JIS. Similarly, for the CG, we had to iteratively reset the bandwidth parameters so as to achieve optimal performance, which is also not feasible in real situations.

The treecode method failed to converge in three cases, in which the brain region overlap between image pairs in the trial data set was less than 50%. In these cases, accurate modeling of the density function reveals the off-diagonal JIS clusters, which then causes the registration process to diverge. The average registration error was slightly worse compared with the other two tested methods owing to the random oscillations of the motion parameters about the gold-standard position in the original resolution scale. This issue was observed before for satellite images and is due to the random sampling scheme.

Second, the proposed method can register a maximum of ten images at once due to the implementation of the hierarchical subdivision scheme (see Section III-D). As the hierarchical tree is constructed, each cell containing more than two separate points is recursively subdivided into 2^D subcells down

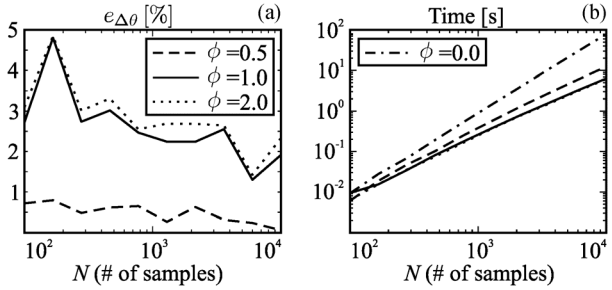


Fig. 7. Treecode properties: (a) relative error $e_{\Delta\theta}$ of the approximated motion update $\Delta\theta$ and (b) corresponding computation times shown for different cell acceptance threshold ϕ and w.r.t. the number of samples N. Note that, in (b), the curve for $\phi = 0.0$ represents times measured for direct implementation.

to the maximal tree level $L_{\max}$ (set to 32). Similarly to [4], only the nonempty cells are stored and tagged by expressions in (15)–(18); however, the problem lies in the indexing scheme needed to reach the cells, which has exponential storage complexity of $\mathcal{O}(2^{L_{\max}D})$. Note that this issue is related only to the implementation and can be therefore avoided by using a more advanced indexing scheme, e.g., hash functions, or other space partitioning methods.

E. Nonrigid Registration

Results for the BrainWeb nonrigid registration are shown in Table I and the corresponding box–whisker diagram in Fig. 6(e). Initial average registration error was 9.5 pixels. The average registration errors for the successful trials were 0.595, 0.621, and 0.119 pixels, respectively, for the EC and the proposed treecode methods based on the Gaussian kernel and the Hilbert kernel. The treecode method based on the Hilbert kernel successfully registered all trials and image pairs, whereas the EC and the Gaussian-based treecode achieved 88% (91.3%) and 79% (85.7%) trial (image pair) success rates. Both the EC and the Gaussian-based treecode methods successfully aligned the skin-background edges, whereas they had difficulties registering the brain structures that lack strong edge information. Thus, most of the registration error resulted from the misaligned central control point (c.f. Fig. 5). The average times for a trial image group to register were 127 and 107 s for the EC and for the proposed treecode methods, respectively.

F. Properties and Performances of the Treecode Method

Fig. 7 shows the relative error $e_{\Delta\theta}$ of the approximated motion update $\Delta\theta^*$ and corresponding computation times shown for different cell acceptance thresholds ϕ and w.r.t. the number of samples N. The graphs were obtained by repeatedly drawing samples of different sizes N from one trial data set so as to factor out the influence of the data distribution. For $\phi = 0.5$, the relative error is strictly below 1%, whereas for $\phi \geq 1$, it is about 3%. Note that the relative error reduces gradually for higher N.

For $\phi \geq 1$, the approximate motion update $\Delta\theta^*$ computed from (19) and (20) is obtained in shorter times ($< 1/10$) than by direct implementation of (13) for each of the tested samples sizes N. Execution times are longer for $\phi = 0.5$ since the tree-structured hierarchical JIS subdivision is explored down to smaller cells so as to satisfy the cell acceptance threshold. Compared with direct implementation, the actual speedup is still significant, particularly for higher N.

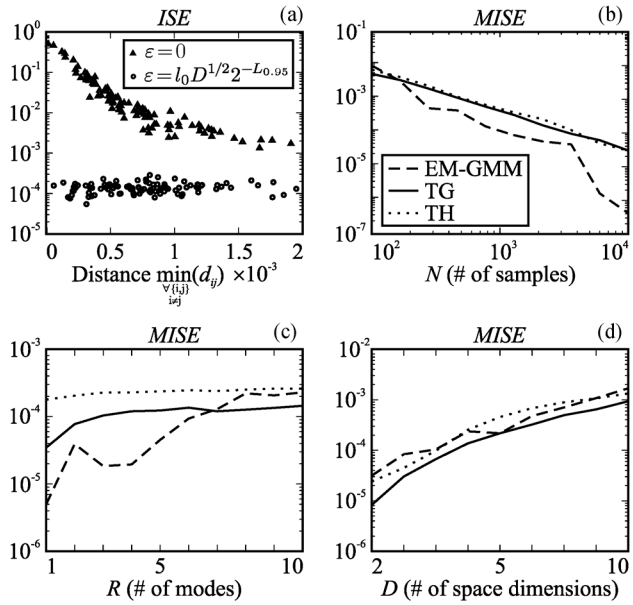


Fig. 8. Density estimation performances according to ISE (c.f. Section IV-F). Effect of Hilbert kernel regularization ε is shown in (a), while (b), (c), and (d) plot the dependence of MISE w.r.t. N, R, and D, respectively. In all tests, ten repetitions were run to obtain MISE. Default testing parameters were $N = 2000$, $R = 5$, and $D = 5$, except when observed as the independent variable. EM-GMM was run with $R^* = 5$ components.

The dependence of ISE on the observed minimal distance between the intensity points in JIS is shown in Fig. 8(a). By setting ε according to (22), ISE is improved and becomes independent of the minimal intensity point distances.

In the following, we study the performance of density estimators, applied in the image registration experiments, in terms of MISE for N samples drawn from a known density function $p(z)$, defined in (25), with R modes in D-dimensional space. As expected, the performance of the density estimators increases by drawing more samples from $p(z)$. Conversely, the performance decreases with increasing the dimensionality of the underlying true density function [c.f. Fig. 8(c) and (d)].

The most interesting results are for varying R in the true density function. The treecode-based density estimators exhibit steady performance with MISE $\sim 10^{-4}$, whereas the EM-GMM method has superior performance for $R^* > R$ [see Fig. 8(c)]. For $R^* < R$ the performance of EM-GMM drops significantly since the separate modes in the density function are modeled using fewer insufficient number of GMM components. This observations further stress the issue of choosing the correct number of GMM components R^* for the EC registration method. Although, for the experiments in Sections V-B–V-D, the EC was run with higher R^* than that suggested in [5], the results obtained were significantly worse. This indicates that, even for $R^* \geq R$, the EM-GMM does not necessarily find all modes of the density function (see Fig. 9), i.e., the information which is not reflected in MISE but plays a crucial role in the image registration process.

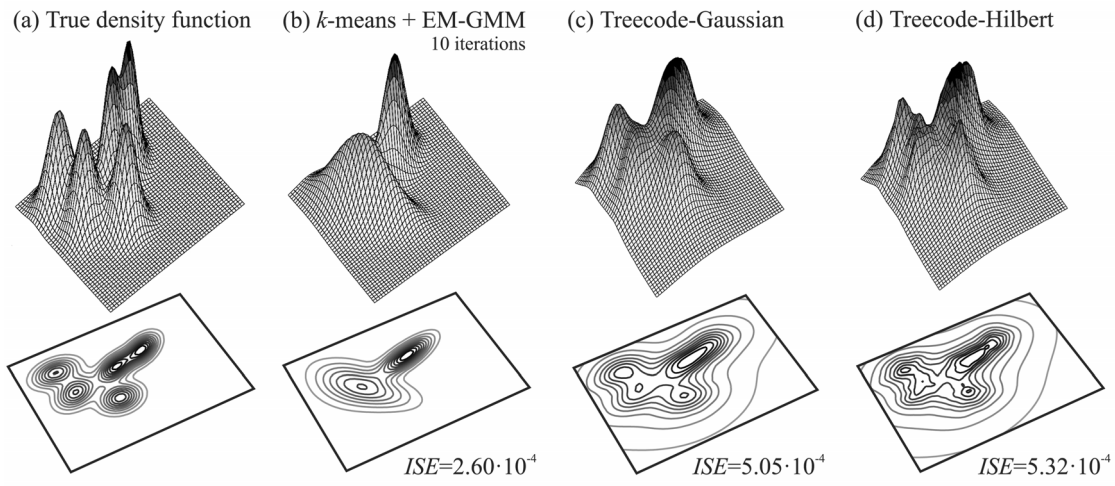


Fig. 9. Artificial density function in (a) has five distinct modes. In (b), the EM-GMM estimate had the same number of components as the true number of modes but found only two modes; in (c), the treecode with the Gaussian kernel found four; and in (d), with the Hilbert kernel, it found all five modes. Comparing the pointwise accuracy of density estimates by ISE indicates that ISE puts more stress on the tails of the density and does not reflect the missed modes.

VI. Discussion

Results indicate that the proposed method has similar and, on two data sets, even superior performances compared with the EC method [5] in terms of the registration error, the trial success rate, and the image-pair success rate. Both methods share the Newton-type minimization scheme so as to avoid biasing the results. Global criteria are also similar since, under certain assumptions, maximizing log-likelihood can be shown to be equal to minimizing the joint entropy [3]. The main difference between the methods is in the estimation of the JDF. In the EC method, a parametric GMM is used, where the number of mixture components that correspond to the number of modes in the JDF has to be determined in advance either manually or automatically. As shown in the results section, this parameter critically defines the performances of the EC method. Moreover, the actual number of modes in the JDF can vary during the registration process, which may also require adapting the number of components. Instead of modeling explicitly the modes, the proposed method implements a Parzen kernel density method through the hierarchical JIS subdivision scheme to obtain a representative high-dimensional JDF estimate.

By using the Hilbert kernel in the derivation of the joint entropy, scalar terms $W_{ij}^{\theta}\,\|d_{ij}^{\theta}\|^{-2}$ in (13) reveal close resemblance to the gravitational potential $g(r)/r^2$. This observation inspired the hierarchical intensity-space subdivision method [25] that enabled an approximate (relative error $<3\%$) but highly efficient computation ($10\times$ speedup) of the motion parameters $\Delta\theta^*$. We have demonstrated through various image registration and MISE-performance experiments that the Hilbert kernel provides sufficiently accurate JDF estimates, which reveal a higher number of true modes of the underlying density function, compared with the frequently oversmoothed density estimates obtained by methods using the Gaussian kernel. Finding the modes of the density functions is essential to define the correct correspondences between the multimodal images that guide the registration process.

The computational complexity of the proposed method is $\mathcal{O}(N^* \log N^* P^2 D^3)$, which is higher than $\mathcal{O}(\mathrm{RNP}^2 D^3)$ of the EC method. However, a number of improvements can be made to reduce the computational complexity of the proposed method. First, the hierarchical JIS subdivision could be repeated only over a fixed number of motion updates, similarly to the update of the GMM in the EC method. Second, the approximation scheme for the joint entropy currently computes only the point-to-group interactions in JIS, whereas a significant speed-up can be achieved if the interactions would also be computed between groups of intensity points. This should cut the computational complexity down to $\mathcal{O}(NP^2D^3)$, which we plan to explore in our future work.

If a large number of images is used in the registration process, then resolving the intensity correspondences from JIS may be hampered by the so-called *curse of dimensionality*. The image overlap region defines the number of intensity points N used to estimate the JDF. However, the dimensionality of the JDF equals the number of images D used in the registration process. Therefore, as more images are added to the registration process, the JIS will be increasingly more sparse, and the relative distances between the intensity points will become more similar and thus less informative. In other words, the clusters in JIS will be harder to detect, and only the intensity points closest in the distance will guide the alignment of the images. Therefore, for extremely large image data sets, both the EC and the proposed method will not be robust. We have demonstrated through experiments that the proposed method is capable of registering up to ten images simultaneously, i.e., the density function and its gradient were effectively computed from a 10-D JIS.

Groupwise registration methods typically employ clustering techniques to efficiently represent the intensity points in JIS [5], [17], [18]. The proposed method inherently performs clustering on the cooccurring intensity pairs via the hierarchical intensity-space subdivision method. However, clustering is adapted via the cell acceptance threshold (ϕ) to efficiently compute each partial contribution to the joint entropy. Entropic graph methods

[19], [20] perform clustering based on the edge length between intensity points, which was shown to be a consistent estimator of α entropy [19]. In the proposed method, the Hilbert kernel estimate of the JDF is a harmonically weighted nearest-neighbor JDF estimate [24]. Therefore, the proposed method can be seen as an adaptive k-nearest neighbor entropic graph method.

VII. Conclusion

We proposed a method for groupwise registration of multimodal images that has no data-dependent tuning parameters and is fully automatic. The registration performances are comparable to the state-of-the-art EC method, under the condition that the EC is initialized with an optimal number of Gaussian components for the GMM. However, this parameter might be hard to determine in practical image registration applications. On the other hand, the proposed method uses the nonparametric approach to JDF estimation that automatically captures all the important features needed to detect the correct intensity correspondences between multimodal images that are being registered.

In addition to the rigid-body, affine, and TPS transformation models, the proposed treecode method supports any class of parametric spatial transformations. To use the nonparametric registration framework, the joint entropy could be minimized directly, and the method would also need to incorporate a transformation regularization term.

Further research is directed toward extending the hierarchical intensity-space subdivision scheme so as to enable an efficient computation of the multivariate extensions of mutual information [4]. In our opinion, additional entropy terms would favorably reward image complexity in the overlapping image domain and thereby improve registration performances.

Acknowledgment

The authors would like to thank the following organizations for the use of their data: the RIRE project, the BrainWeb project, Landsat, and the Yale Face Database. Medical Image Analysis Lab is acknowledged for providing the fMRI Simulation Toolbox (SimTB).

References

[1] A. Guimond, A. Roche, N. Ayache, and J. Meunier, "Three-dimensional multimodal brain warping using the Demons algorithm and adaptive intensity corrections," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 58–69, Jan. 2001.

[2] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–98, Apr. 1997.

[3] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, Sep. 1997.

[4] J. Zhang and A. Rangarajan, "Multimodality image registration using an extensible information metric and high dimensional histogramming," *Inf. Process. Med. Imag. LCNS*, vol. 3565, pp. 725–737, 2005.

[5] J. Orchard and R. Mann, "Registering a multisensor ensemble of images," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1236–1247, May 2010.

[6] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[7] H. Park, P. H. Bland, A. O. Hero, and C. R. Meyer, "Least biased target selection in probabilistic atlas construction," *Med. Image Comput. Comput.-Assisted Intervention, LNCS*, vol. 3750, pp. 419–426, 2005.

[8] S. Marsland, C. Twining, and C. Taylor, "A minimum description length objective function for groupwise non-rigid image registration," *Image Vis. Comput.*, vol. 26, no. 3, pp. 333–346, Mar. 2008.

[9] T. F. Cootes, C. J. Twining, V. S. Petrovi, K. O. Babalola, and C. J. Taylor, "Computing accurate correspondences across groups of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1994–2005, Nov. 2010.

[10] Q. Wang, L. Chen, P.-T. Yap, G. Wu, and D. Shen, "Groupwise registration based on hierarchical image clustering and atlas synthesis," *Hum. Brain Mapp.*, vol. 31, no. 8, pp. 1128–1140, Aug. 2010.

[11] G. Wu, H. Jia, Q. Wang, and D. Shen, "SharpMean: Groupwise registration guided by sharp mean image and tree-based registration," *NeuroImage*, vol. 56, no. 4, pp. 1968–1981, Jun. 2011.

[12] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta, "Automated image registration: I. General methods and intrasubject, intramodality validation," *J. Comput. Assist. Tomogr.*, vol. 22, no. 1, pp. 139–152, Jan. 1998.

[13] E. G. Learned-Miller, "Data driven image models through continuous joint alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 236–250, Feb. 2006.

[14] L. Zöllei, "A unified information theoretic framework for pair- and group-wise registration of medical images," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, 2006.

[15] Q. Wang, G. Wu, P.-T. Yap, and D. Shen, "Attribute vector guided groupwise registration," *NeuroImage*, vol. 50, no. 4, pp. 1485–1496, May 2010.

[16] G. Wu, Q. Wang, H. Jia, and D. Shen, "Feature-based groupwise registration by hierarchical anatomical correspondence detection," *Hum. Brain Mapp.*, vol. 33, no. 22, pp. 253–271, Feb. 2012.

[17] C. Studholme and V. Cardenas, "A template free approach to volumetric spatial normalization of brain anatomy," *Pattern Recognit. Lett.*, vol. 25, no. 10, pp. 1191–1202, Jul. 2004.

[18] J. Orchard and L. Jonchery, "Ensemble registration: Aligning many multi-sensor images simultaneously," *Proc. SPIE*, vol. 33, pp. 724 50O-1–724 50O-12, 2009.

[19] B. Ma, A. Hero, J. Gorman, and O. Michel, "Image registration with minimum spanning tree algorithm," in *Proc. IEEE Int. Conf. Image Process.*, Vancouver, BC, Canada, 2000, vol. 1, pp. 481–484.

[20] H. Neemuchwala, A. Hero, S. Zabuawala, and P. Carson, "Image registration methods in high-dimensional space," *Int. J. Imag. Syst. Technol.*, vol. 16, no. 5, pp. 130–145, 2006.

[21] S. Watanabe, "Information theoretical analysis of multivariate correlation," *IBM J. Res. Develop.*, vol. 4, no. 1, p. 66, 1960.

[22] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, 1962.

[23] S. J. Sheather, "Density estimation," *Stat. Sci.*, vol. 19, no. 4, pp. 588–597, Nov. 2004.

[24] L. Devroye and A. Krzyzak, "On the Hilbert kernel density estimate," *Statist. Probab. Lett.*, vol. 44, no. 3, pp. 299–308, Sep. 1999.

[25] J. Barnes and P. Hut, "A hierarchical $\mathcal{O}(N \log N)$ force-calculation algorithm," *Nature*, vol. 324, no. 6096, pp. 446–449, 1986.

[26] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, Jr., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. Hill, D. J. Hawkes, C. Studholme, J. B. Maintz, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, G. Q. Maguire, Jr., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *J. Comput. Assist. Tomogr.*, vol. 21, no. 4, pp. 554–566, Jul./Aug. 1997.

[27] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

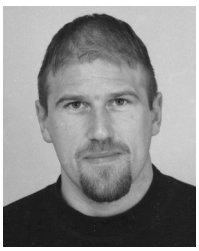
[28] R. K. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1085–1097, Nov. 1999.

[29] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.

[30] H. Y. Kim and J. Orchard, "Registering a non-rigid multi-sensor ensemble of images," in *Proc. IEEE Eng. Med. Biol. Conf.*, Buenos Aires, Argentina, 2010, vol. 1, pp. 5935–5938.

Žiga Špiclin received the B.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, in 2006 and 2011, respectively.

He is currently a Research Fellow with the Laboratory of Imaging Technology, Faculty of Electrical Engineering, University of Ljubljana. His research interests include computer vision and biomedical and hyperspectral imaging, with particular interest for the development of image registration, restoration, and reconstruction techniques.



Boštjan Likar received the B.Sc., M.Sc., and Ph.D. degrees from the University of Ljubljana, Ljubljana, Slovenia, in 1995, 1998, and 2000, respectively, all in electrical engineering, and the Ph.D. degree from Utrecht University, Utrecht, The Netherlands, in 2000.

Since 1996, he has been with the Faculty of Electrical Engineering, University of Ljubljana, where he is currently a Full Professor. He is the author of over 60 SCI journal papers with over 400 clear citations, a reviewer of over ten SCI journals, a principal researcher of six research projects, a designer of more than ten new computer vision products, and a cofounder of the high-tech company Sensum. His research interests concentrate on visual quality inspection, computer and machine vision systems, and biomedical and hyperspectral imaging.

Dr. Likar is a program committee member of over 15 international conferences.

Franjo Pernuš received the Diploma, M.S., and Ph.D. degrees in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, in 1976, 1979, and 1991, respectively.

Since 1976, he has been with the Department of Electrical Engineering, University of Ljubljana, where he is currently a Professor and the Head of the Imaging Technologies Laboratory. He is a Cofounder of Sensum, a company that supplies machine vision solutions for the pharmaceutical industry. He is an author or coauthor of over 150 refereed scientific articles on biomedical image processing and computer vision and has supervised seven Ph.D. students. His research interests include biomedical image processing and analysis, computer vision, and the applications of image processing and analysis techniques to various biomedical and industrial problems.

Dr. Pernuš is an Associate Editor of the IEEE TRANSACTIONS ON MEDICAL IMAGING and of *Computer Aided Surgery*. In the past, he was an Associate Editor of *Pattern Recognition Letters* and of *Electrotechnical Review*.