Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Medical Image Analysis 7 (2003) 539–552

MEDICAL IMAGE ANALYSIS

www.elsevier.com/locate/media

MAP MRF joint segmentation and registration of medical images

Paul P. Wyatt*, J. Alison Noble

Medical Vision Laboratory, Department of Engineering Science, University of Oxford, Oxford, UK

Abstract

The problems of segmentation and registration are traditionally approached individually, yet the accuracy of one is of great importance in influencing the success of the other. In this paper, we aim to show that more accurate and robust results may be obtained through seeking a joint solution to these linked processes. The outlined approach applies Markov random fields in the solution of a *maximum a posteriori* model of segmentation and registration. The approach is applied to synthetic and real MRI data.

Keywords: Segmentation; Registration; Joint; Combined

1. Introduction

Two of the most fundamental problems in medical image analysis are those of segmentation: meaningful labelling of raw data, and registration: alignment of information from multiple datasets. In most previous work segmentation and registration have been performed consecutively which has the disadvantage that errors propagate from one to the other. This paper examines whether combining segmentation and registration yields any advantage in accuracy, speed or robustness.

The major challenge in combining segmentation and registration is to ensure convergence and prevent poor estimates of either segmentation or registration from harming the other. A key idea is to incorporate partially registered datasets in a combined class model and estimates of segmentation labels in a registration criteria. Theoretically this employs the total available information more advantageously and benefits both classification and registration accuracy. An additional challenge is to avoid a significant increase in the computational load.

Segmentation, in the sense of spatial clustering, attempts to reduce the variation in image appearance to a small set of discrete labels. The accuracy of this process is dependent on the presence of noise, intensity inhomogeneities and biological factors. In the absence of an application-specific prior model, the success of any approach is dependent primarily on the discriminative power of the likelihood model for the data. It was demonstrated by Hurn et al. (1996) that multiple measurements could improve segmentation for preregistered data. This observation is true for multispectral imaging (Clark et al., 1998) and for mono/multimodal sequences where the imaged object is in motion as for many imaging modalities this also decorrelates noise. It is reasonable to expect an increase in accuracy from combined segmentation over separate segmentation if errors arising from misalignment can be contained.

Registration attempts to obtain a transformation which matches multiple data sets, under the assumption that some correlation exists. Current methods are divisible into two types; pixel-based and feature-based techniques (Brown, 1992; Maintz and Viergever, 1998). For different reasons both approaches suffer problems. Pixel-based methods rely on describing similarity between the images as some distance measure calculated from the pixel (intensity) values. The best candidate point matches are found and regularized under the general assumption that the deformation field is continuous and smooth. This assumption tends to result in boundary delocalization as neighbouring regions often have different physical properties. This is difficult to avoid without knowledge of class. Feature-based methods are those where matches between (anatomically) significant points are sought. Biological structures are often difficult to mark with repeatable accuracy and as such introduce localization error. This is

*Corresponding author.
E-mail addresses: wyatt@robots.ox.ac.uk (P.P. Wyatt), noble@robots.ox.ac.uk (J.A. Noble).

1361-8415/03/$ – see front matter
doi:10.1016/S1361-8415(03)00067-7

compounded where geometric information is not retained and only the point's location is used (Pennec, 1998). In addition, image features are typically extracted semi-automatically and the time intensiveness for the user leads to few points being used and a sparse field approximation to the true deformation. The addition of a segmentation label to registration potentially allows these problems, for both types of registration, to be overcome. The maximum a posteriori estimate of a segmentation label field is essentially a probabilistic edge map, in addition to being a region segmentation field. Retaining a probabilistic description of edges as opposed to a binary edge map should create a more discriminative and robust registration measure. It also provides knowledge of where the deformation is not likely to be smooth for non-rigid registration.

In seeking to combine segmentation and registration it is desirable to avoid the mentioned difficulties. Additionally consecutive implementation can lead to propagation of errors from one to the other. This is difficult to detect and impossible to correct without their combination. In theory, combination should produce two principal advantages: greater accuracy and robustness. In this paper a combined algorithm is developed to demonstrate these points.

Combined segmentation and registration has potential applications in a number of areas; essentially wherever both processes have previously been performed sequentially. Examples include the registration of skeletonized arterial trees in MR angiography (McLaughlin et al., 2002), tracking the myocardium in cardiac imaging and treatment of tumours in breast imaging.

There is little previous work in this area. Hurn et al. (1996) noted that registered images could be fused using a Bayesian MRF approach to obtain greater accuracy than was possible separately. Bansal et al. (1999) developed a Min–Max entropy-based algorithm for registration of 2D X-ray portal and 3D CT images. Their algorithm aimed to estimate the segmentation and registration of the portal images with respect to the known CT data. They assumed that the segmentation labels are pixel-wise independent and so essentially obtained the maximum likelihood segmentation. The method switched between estimating the maximum entropy segmentation and minimum entropy registration. The registration was based upon a mutual information criterion (Maes et al., 1997; Wells et al., 1996), modified to weight the joint histogram by the likelihood of the segmentation labels.

An active contour approach was proposed by Yezzi et al. (2001) where registration was obtained from contour propagation. The contour was defined by an energy function dependent upon all the images. The segmentations were obtained from the final contour position; related to one another via the registration. The solution was obtained through a two-step gradient-descent algorithm. This algorithm is suited to tasks where only a point-to-point correspondence between single contours is required, but suffers the drawback of being highly initialization-dependent when multiple objects exist. In addition, only a single contour of point-to-point correspondences is obtained.

An alternative to general combined methods are application-specific Active Appearance Models and Atlas techniques (Cootes and Taylor, 2001; Pizer et al., 1999). In these a model is built through a principal component analysis of a training set of examples. Although theoretically very effective, good performance requires training data which captures all significant variation. Also, similarly to active contours, they can be highly initialization-dependent.

We favour a Markov random field framework, a mathematical technique for embedding local spatial information (Besag, 1974, 1986; Geman and Geman, 1984), within which we seek to obtain a *maximum a posteriori* estimate of the segmentation and registration. This aids implicit embedding of spatial information into the registration criterion. This paper primarily focuses on rigid registration, though extension to non-rigid registration and examples of its application are also considered in Sections 3.3 and 5. In Section 2, a Bayesian description of combined segmentation and registration is presented. Section 3 describes the image models used for segmentation and registration, before the general algorithm and details of implementation are examined in Section 4. Results on simulated and real data are shown in Section 5 and discussed in Section 6.

2. Integrating segmentation and registration

Our goal is to *obtain the best possible estimate, in some predefined sense, for the segmentations of n multiple data sets degraded by non-stationary noise, which are related through some geometric transformation, and to recover the geometric transformation.* We cast this as a *maximum a posteriori* (MAP) estimation of the segmentation labels $\mathcal{S} = [S_1, S_2, \ldots, S_n]$, transformation(s) $\mathcal{T}$ given n datasets $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n$ and pose the solution using Markov Random Fields (MRFs).

Use of MRFs and Bayesian MAP estimation requires a model to be defined for the segmentation and registration processes, conditioned upon the data. This is important, as no matter how detailed the prior information available for a given class of problems, the data determines a specific instance of that problem. The choice of a prior model, an image description with parameter set θ_k, $k \in \mathcal{K}$, and Markov spatial relationship, is also critical as it determines the expected relationship between data and class.

The Bayesian problem may be stated for two datasets 1 and 2 as

$$\mathcal{P}(\mathcal{S}, \mathcal{T}/\mathbf{X}_1, \mathbf{X}_2) = \frac{\mathcal{P}(\mathbf{X}_1, \mathbf{X}_2/\mathcal{S}, \mathcal{T})\mathcal{P}(\mathcal{S}, \mathcal{T})}{\mathcal{P}(\mathbf{X}_1, \mathbf{X}_2)}. \quad (1)$$

If data independence is assumed, then $\mathcal{P}(\mathbf{X}_1, \mathbf{X}_2) =$

$\mathcal{P}(\mathbf{X}_1)\mathcal{P}(\mathbf{X}_2)$ which, for consistency, implies $\mathcal{P}(S_1, S_2, \mathcal{T}) = \mathcal{P}(S_1)\mathcal{P}(S_2)\mathcal{P}(\mathcal{T})$. This assumption is used to initialize the registration as it leads to writing

$$\mathcal{P}(S_1, S_2, \mathcal{T}/\mathbf{X}_1, \mathbf{X}_2) = \frac{\mathcal{P}(\mathbf{X}_1/S_1)\mathcal{P}(S_1)\mathcal{P}(\mathbf{X}_2/S_2)\mathcal{P}(S_2)\mathcal{P}(\mathcal{T})}{\mathcal{P}(\mathbf{X}_1)\mathcal{P}(\mathbf{X}_2)}. \quad (2)$$

Experiments have shown that it is more robust to initialize the parameters θ for each dataset individually as if a poor registration exists initially they can become prone to error. The proposed algorithm switches between Eqs. (1) and (2) as their validity changes. Typically it switches once, from using Eq. (2) to using Eq. (1). Details are provided in Section 3.2.

In the proposed method Eqs. (1) and (2) are not implemented directly. Instead we expand $\mathcal{P}(\mathcal{S}, \mathcal{T})$ using Bayes' rule and take the logarithms of both sides. The denominator is dropped as it is constant with respect to the data and optimization:

$$\ln \mathcal{P}(\mathcal{S}, \mathcal{T}/\mathbf{X}_1, \mathbf{X}_2) \propto \ln \mathcal{P}(\mathbf{X}_1, \mathbf{X}_2/\mathcal{S}, \mathcal{T}) + \ln \mathcal{P}(\mathcal{S}/\mathcal{T}) + \ln \mathcal{P}(\mathcal{T}). \quad (3)$$

For rigid registration, $\mathcal{P}(\mathcal{T})$ might also be dropped as it is typically difficult to estimate a sensible prior on the global transform other than a broad range of limits. For non-rigid registration, $\mathcal{P}(\mathcal{T})$ is the deformation model imposed upon the transformation.

Throughout this paper, we refer primarily to joint estimation of segmentation and registration. It is therefore necessary to distinguish it from simultaneous estimation. We define **simultaneous** estimation as *updating the estimation of both the classes and transforms relating any* (*two*) *datasets in a single step optimization*. **Joint** estimation may, or may not, use the same model of segmentation and registration, but alternates between *updating the classes of any* (*two*) *datasets and updating the geometric transforms between them in a multi-step* (*two or more steps*) *optimization*.

To illustrate this, with respect to Eq. (3), the joint scheme would be

$$\ln \mathcal{P}(\mathcal{S}_{n+1}, \mathcal{T}_n/\mathbf{X}_1, \mathbf{X}_2) \propto \mathcal{P}(\mathbf{X}_1, \mathbf{X}_2/\mathcal{S}_n, \mathcal{T}_n) + \ln \mathcal{P}(\mathcal{S}_n/\mathcal{T}_n) + \ln \mathcal{P}(\mathcal{T}_n),$$
$$\ln \mathcal{P}(\mathcal{S}_{n+1}, \mathcal{T}_{n+1}/\mathbf{X}_1, \mathbf{X}_2) \propto \ln \mathcal{P}(\mathbf{X}_1, \mathbf{X}_2/\mathcal{S}_{n+1}, \mathcal{T}_n) + \ln \mathcal{P}(\mathcal{S}_{n+1}/\mathcal{T}_n) + \ln \mathcal{P}(\mathcal{T}_n). \quad (4)$$

Note that segmentation is performed first in order that the class labels are available for the subsequent registration. The order is important for robustness as it ensures that spatial regularization is used from the start. If it is desired to perform registration first then initial labels from, for example, the K-Means algorithm (Bishop, 1995) could be used.

The simultaneous estimation is

$$\ln \mathcal{P}(\mathcal{S}_{n+1}, \mathcal{T}_{n+1}/\mathbf{X}_1, \mathbf{X}_2) \propto \ln \mathcal{P}(\mathbf{X}_1, \mathbf{X}_2/\mathcal{S}_n, \mathcal{T}_n) + \ln \mathcal{P}(\mathcal{S}_n/\mathcal{T}_n) + \ln \mathcal{P}(\mathcal{T}_n). \quad (5)$$

The difference between these schemes depends significantly upon the models for segmentation and registration and their interdependence. As they become more independent the differences diminish. As the models become more reliant the gains in simultaneous estimation appear. However, joint estimation is significantly quicker than simultaneous estimation. This results simply from the number of classes to be considered. If there are $\mathcal{K}$ distributions and we allow the registration classes to be $\pm\delta\mathcal{T}$ in each dimension $\mathcal{D}$, then for joint estimation we consider $\mathcal{K} + (2\mathcal{D} + 1)$ classes and for simultaneous estimation $\mathcal{K}(2\mathcal{D} + 1)$, i.e. additive versus multiplicative computational complexity. More importantly, in this paper we principally consider *rigid* registration and consequently the registration criterion is evaluated over the image as a whole. For the joint algorithm, it has been found that minimizing with respect to each direction $(\mathcal{S}, \mathcal{T})$ in turn, via the Powell algorithm, is equally accurate to alternating single steps in each direction. However, it is slightly slower.

3. Segmentation and registration

3.1. Segmentation

In the following sections, a familiarity with the general principles of Markov random fields (MRFs) and Hidden MRF (HMRF) techniques (Besag, 1986; Geman and Geman, 1984; Greig et al., 1989) is assumed. An HMRF differs from an MRF in that it is assumed for the MRF that the parameters θ from which the image is generated are known a priori whereas the HMRF iteratively updates an estimate of θ concurrently with estimating the field labels. In practice, an MRF parameter set θ is normally calculated from a K-Means or EM algorithm (Bishop, 1995).

It is assumed that the observed image is a rectangular $2\mathcal{D}$ Gaussian Markov random field, taking labels $q \in \mathcal{Q}$. The MRF used to model the local lattice interaction assumes first order nearest neighbour cliques; each site has four neighbours. The prior density $\mathcal{P}(\mathcal{S})$ is modelled using the common Ising model:

$$\ln \mathcal{P}(\mathcal{S}) \cong \sum_{c \in \mathcal{C}} \beta \mathcal{V}_c(\mathcal{S}) = \sum_{\langle q,r \rangle} \beta(1 - \delta(s_q, s_r)), \quad (6)$$

where $\mathcal{V}_c$ are clique potentials and s_q, s_r are lattice sites forming the nearest neighbour cliques. The clique potential is the Kronecker delta function δ and $\beta > 0$ is a control parameter weighting belief in the prior versus the data likelihood. The sites are isotropic and homogeneous; meaning they are independent of orientation and position on the lattice.

A Gaussian mixture model (GMM) is used to model the likelihood of the image data at a site, x_q, conditioned on the class parameters $\theta_k = \{\mu_k, \sigma_k, \omega_k\}$, $k \in \mathcal{K}$, which are the mean, standard deviation and weight of the distribution:

$$\mathcal{P}(x_q/s) = \prod_{k\in\mathcal{K}} \frac{\omega_k}{(2\pi)^{1/2}\sigma_k} \exp\left\{ -\frac{1}{2}\left(\frac{x_q - \mu_k}{\sigma_k}\right)^2 \right\}. \quad (7)$$

When the joint class models are formed, Eq. (7) is replaced by its multidimensional equivalent. The GMM is ideal for illustrating the benefit of our approach compared to sequential segmentation and registration. Alternative models could be substituted.

The posterior labels are estimated from the GMM likelihood model (Eq. (7)) and the MRF spatial prior (Eq. (6)) which are combined yielding

$$\ln \mathcal{P}(s/x_q) \propto \sum_{k\in\mathcal{K}} \left\{ \ln(\omega_k) - \ln(\sigma_k) - \frac{1}{2}\left(\frac{x_q - \mu_k}{\sqrt{2\sigma_k}}\right)^2 \right\} + \sum_{\langle q,r\rangle} \beta \mathcal{V}_c(s). \quad (8)$$

We use the iterated conditional modes (ICM) algorithm (Besag, 1986) to minimize Eq. (8). The ICM algorithm is chosen to ensure quick convergence to a local minimum (~4–8 iterations). Let $\hat{\mathbf{s}}$ represent class labels and $\mathbf{x}$ data. We seek, at each step,

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{s}\in\mathcal{S}} \{\mathcal{P}(\mathbf{x}|\mathbf{s})\mathcal{P}(\mathbf{s})\}. \quad (9)$$

Initially, before any registration is performed, the Gaussians are of single dimension ($1\mathcal{D}$) and model the separate image intensities. When sufficiently well registered, which is determined as described in Section 3.2, combined 2D Gaussian models are formed using the intensity information from both datasets. Combination takes the parameter set for each Gaussian to $\theta_k = \{\mu_{k1}, \mu_{k2}, \sigma_{k1}, \sigma_{k2}, \sigma_{k3}, \omega_k\}$, $k \in \mathcal{K}$. It is this combination of data which increases the separation of distribution centres and reduces classification error, see Appendix A.

When we switch from separate to combined class models, the new distribution parameters must be determined. In both (Maes et al., 1997) and (Wells et al., 1996), the principle of increasing information of intensity histograms is used to align images. The same principle can be used to determine which 1D segmentation classes match and combine to form the 2D classes. The joint *intensity* histogram is replaced with a joint *class* histogram. Fig. 1 shows a diagram of the joint class histogram as it might look during registration. As the entropy increases, single matrix elements will be both row and column dominant. This dominance identifies the most likely candidates for matching. The joint class histogram is the matrix $\mathcal{P}(\mathcal{S}/\mathcal{T})$, from Eq. (1), and provides the weights ω_k for the combined distributions.

$$\begin{array}{c|ccc} & A_1 & A_2 & A_3 \\ B_1 & 0.01 & 0.05 & 0.24 \\ B_2 & 0.33 & 0.01 & 0.03 \\ B_3 & 0.01 & 0.20 & 0.13 \end{array} \Rightarrow \begin{array}{c|ccc} & A_1 & A_2 & A_3 \\ B_1 & N & R & D \\ B_2 & D & N & R \\ B_3 & N & D & R \end{array}$$

N:Null, D:Dominant, R:Retain ($\omega_l \geq 0.03$)

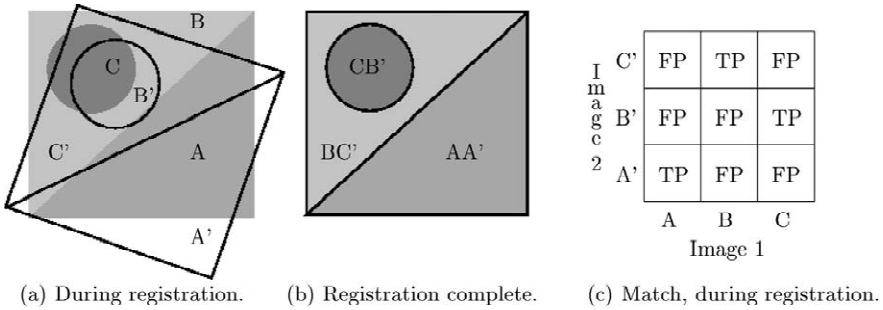
Fig. 2. Obtaining weights ω_l from the joint class histogram dominant elements.

A sample joint class histogram is shown in Fig. 2. On the left are the weight elements ω_l as might be seen at some stage in the registration. Some of these elements are dominant, some are significant enough to retain and the rest may be attributed to error. From the obtained matches at a given point during registration, the class parameters θ_l can be estimated in one of two ways. Firstly the parameters of the 1D distributions can simply be projected into 2D, the weights ω_k being obtained from the joint class histogram, and correlation assumed zero. Note that in combining datasets $\mathbf{X}_i$, $\mathbf{X}_j$ that

$$\mathcal{K}_i\mathcal{K}_j \geqslant \mathcal{K}_{ij} > \max(\mathcal{K}_i, \mathcal{K}_j). \quad (10)$$

The combined number of classes must be greater than the maximum of the separate to allow for errors in registration. Otherwise, if diagonalization is enforced then subsequent registration will be in error.

Alternatively, expectation maximization (EM) estimation can be employed. The combined parameters are then iteratively re-estimated from the current parameters and the



(a) During registration. (b) Registration complete. (c) Match, during registration.

Fig. 1. Diagram of the joint class histogram obtained during registration (TP: true positive; FP: false positive).

combined dataset at the current transformation; $\mathcal{P}(\theta_l^n / \theta_l^{n-1}\mathbf{X}_A, \mathbf{X}_B^{T^n})$. Iterative re-estimation of the parameters turns the MRF into a Hidden Markov random field (Zhang et al., 2001). If re-estimation is employed, then an additional constraint is required to prevent distributions from centering on small error regions in the joint histogram. The constraint is that the model parameters μ_{ij}, σ_{ij} are not independent. As combined distributions have been formed according to Eq. (10), $\mathcal{K}_{i,j=c}$, $c \in \mathcal{K}_j$, share means and variances. Re-estimation of the parameters should enforce that if j is constant then $\mu_{i,j=c} = [\mu_i, \mu_j] = [\mu_i, \text{constant}]$ $\forall i$.

As registration proceeds, the false positives in Fig. 1 tend towards a residual dependent upon classification errors. The relative weights reflect belief in different class combinations, which could be fixed if known for a particular application a priori. Depending on whether we believe a 1:1 correspondence exists between classes in the different datasets; we can allow either a greater or smaller number of classes to be included in the GMM through modifying the weights obtained from the joint class histogram. This allows flexibility in modelling registration as either pure deformation (one-to-one correspondence) or as change between the classes in different images (many-to-one). For example, if an object doubles in size this could be modelled with a diagonal class matrix indicating a unique match between classes and that (say) class A has grown where class B has shrunk. Equally, a correspondence modelled with a non-diagonal class matrix would indicate that part of class A has become B. In terms of application, the former is suited to tracking where the latter is suited to tumour growth or recession.

There are two further roles for the joint class histogram. The first and most important is as a registration criterion and the second is in determining the degree of registration convergence.

3.2. Rigid registration

A class-based information theoretic criteria similar to that of mutual information (Maintz and Viergever, 1998; Roche et al., 1999) is used. At each step of the joint segmentation and registration (JS&R) estimation a sub-optimal set of segmentation labels exists. Consider the following measure:

$$\mathcal{I}_{\mathcal{S}_A, \mathcal{S}_B} = \sum_{k \in \mathcal{K}} \mathcal{P}_k(\mathcal{S}_A, \mathcal{S}_B) \ln \mathcal{P}_k(\mathcal{S}_A, \mathcal{S}_B), \tag{11}$$

calculated from the joint **class** histogram of the two images, instead of the joint **intensity** histogram. This class-based information measure is fast to calculate, as typically 2–5 class labels exist versus (approximately) 256 intensity levels. The smaller matrix size means fewer calculations are required for Eq. (11) compared to mutual information. More importantly, as the class labels of Eq. (11) are those of the posterior estimate they implicitly incorporate spatial information. Mutual information is theoretically similar to maximum likelihood techniques (Roche et al., 1999) and as previously implemented have not incorporated spatial information.

Eq. (11) can be combined with the class pixel MAP probabilities if desired, to increase accuracy, and can be implemented for elastic registration. The labels themselves form a discrete region map and multiplication by their MAP probabilities softens the representation. The MAP probabilities result from the MRF MAP ICM (iterated conditional modes; Besag, 1986) segmentation as calculated from the image data and prior image models.

Using the segmentation labels in the registration criteria allows estimation of the convergence of the registration. The K-Means algorithm allows the estimation of the weightings of each distribution present in the 1D datasets. These weights appear, under maximum likelihood classification, as the row (column) sums of the match matrix of Eq. (11) and are similar under MAP classification. The convergence can be estimated as the ratio of the entropy of the class match matrix, at the nth iteration, to the match matrix where maximal correlation, i.e. a one-to-one correspondence, exists.

$$C^n_{\text{convergence}} = \frac{\sum_{k \in \mathcal{K}} \mathcal{P}_k(\mathcal{S}_A^{\max}, \mathcal{S}_B^{\max}) \ln \mathcal{P}_k(\mathcal{S}_A^{\max}, \mathcal{S}_B^{\max})}{\sum_{k \in \mathcal{K}} \mathcal{P}_k(\mathcal{S}_A^n, \mathcal{S}_B^n) \ln \mathcal{P}_k(\mathcal{S}_A^n, \mathcal{S}_B^n)}. \tag{12}$$

Eq. (12) thereby allows switching between Eqs. (1) and (2), at some user set threshold of convergence. It has been heuristically set at 0.9. This helps avoid obtaining a poor registration which then leads to decreased accuracy in the joint segmentation. It should be noted that the coefficient, in Eq. (12), is nonlinear.

For the purposes of the maximization, $\mathcal{P}(\mathcal{T})$ is dropped for rigid registration as there is normally no useful prior knowledge of the range of the translation and rotation parameters. Rigid registration is performed using the Powell method (Press et al., 1992) and Eq. (11).

3.3. Non-rigid registration

Although we primarily consider segmentation combined with *rigid* registration, a simultaneous block-matching *non-rigid* algorithm has also been implemented. The intention is to demonstrate that, whilst the proposed approach for rigid registration can be extended to non-rigid registration, it is poorly suited for this generalization despite its success on rigid registration.

The principal difference for non-rigid registration is that a transformation regularization prior, $\mathcal{P}(\mathcal{T})$, is required. An isotropic Gaussian prior is assumed for regularization, defining $\mathbf{t}_q = [t_x, t_y]_q$, $\mathbf{t}_r = [t_x, t_y]_r$ as the translations at a pair of neighbouring lattice sites:

$$\mathcal{P}(\mathcal{T}) = \frac{1}{(2\pi)^{1/2}\sigma_t} \exp\left\{ -\frac{1}{2} \left| \frac{\mathbf{t}_q - \mathbf{t}_r}{\sigma_t} \right|^2 \right\}. \quad (13)$$

The parameter σ_t performs, for the transform, the equivalent function to the spatial MRF parameter β.

The non-rigid likelihood registration criteria used is the same as for rigid registration; that of Eq. (11). However, it is updated in a different way. For most, if not all, non-rigid algorithms it is desirable to calculate the likelihood of a point-to-point correspondence *once*. To avoid an expensive recalculation of the likelihoods at each step, it is assumed that they can be calculated *once* from the initial segmentation labels at a given level. Ideally the likelihoods should be recalculated as they are not independent of changes in the segmentation. In our method, a 7×7 block is used to calculate the most likely entropic match over a small area ±4 pixels. Let $\mathcal{I}_{\mathcal{Q}_x}(\mathbf{t})$ denote the class entropy calculated from this 7×7 block subset of the Markov field $\mathcal{Q}$ centred on x and shifted by a local translation $\mathbf{t}$. The likelihood of a particular translation $\mathbf{t}$ is simply $\mathcal{P}(\mathcal{S}_{\mathcal{Q}_x}/\mathbf{t}) = \mathcal{I}_{\mathcal{Q}_x}(\mathbf{t})$.

The combination of these likelihoods with the transform prior of Eq. (13) and segmentation Eq. (8) yields the combined segmentation and registration equation:

$$\ln \mathcal{P}(s, \mathbf{t}/\mathbf{x}_q) \propto \sum_{k\in\mathcal{K}} \left\{ \ln(\sigma_k) - \ln(\omega_k) + \left(\frac{\mathbf{x}_q - \mu_k}{\sqrt{2\sigma_k}} \right)^2 \right\} + \sum_{\langle q,r \rangle} \beta \mathcal{V}_c(s) + \mathcal{I}_{\mathcal{Q}_x}(\mathbf{t}) + \left\{ \ln(\sigma_t) + \left\| \frac{\mathbf{t}_q - \mathbf{t}_r}{\sqrt{2\sigma_t}} \right\|^2 \right\}, \quad (14)$$

where $\mathbf{x}_q$ refers to the joint data $[x_1, x_2^{t_{21}}]$ with $x_2^{t_{21}}$ obtained by nearest neighbour interpolation from the second dataset.

Although block matching is a natural extension of the proposed rigid algorithm, for the purposes of combining segmentation with non-rigid registration it is not ideal. Block matching is hampered by four connected problems relating to computational feasibility, robustness, accuracy and inconsistency in its assumptions.

Firstly, block matching uses a small square of perhaps $\geqslant 7\times7$ pixels to determine matches. It is assumed these are rigid. The block is then shifted to a number of discrete locations and the similarity measure is calculated at each. A reasonable number of points is required in order that a measure such as mutual information has sufficient data to be *relatively* robust. A simultaneous algorithm wishes to update segmentation and registration at each step and for an MRF at each point. This implies a substantial amount of block overlap and hence an extremely high computational cost. Block matching normally calculates a set of candidate likelihood matches which are then regularized. As the segmentation will change at many locations each step, it is not possible to do this in a single step except approximately. The large number of parameters to be estimated for segmentation and non-rigid registration suggest that it be performed simultaneously to avoid local minima. As noted in Section 2 computational complexity of the simultaneous algorithm is significantly higher than the joint algorithm irrespective of additional costs from block matching. If block matching is used and the likelihoods are re-estimated at each step the computational cost is currently prohibitive. The obvious solution of using the pixel intensities, which do not change, loses the benefit of the spatial prior.

The second problem, robustness of block matching, is dependent on the block size compared to the noise level. With noisier modalities larger block sizes tend to be required. This leads to an inconsistency in assumptions: the block used for assessing the candidate likelihoods is discretely displaced by some rigid deformation where the transformation to be recovered is believed non-rigid.

Accuracy is possibly the biggest problem. Block matching takes no inherent account of deformation or rotation of the block. This implies that criteria such as MI or the correlation ratio, which have no inherent invariance to these events, will not cope well with these situations. Additionally, their statistical nature means that information is averaged and hence high frequency changes can be lost. This leads to delocalization in their estimates where regions are not symmetrical, an effect enhanced by the application of an isotropic prior.

To summarize, the poor suitability of an entropy or MI type criterion and block matching for combined segmentation and non-rigid registration results from its requirement for multiple data points. To avoid this requires a basic reformulation of the matching criteria to use point-based information, ideally incorporating some predictive capacity to facilitate the search. The non-rigid algorithm is demonstrated on a sequence of mouse heart images in Section 5.

4. Implementing the joint segmentation and registration algorithm

The complete algorithm, for both the rigid and non-rigid case, is described in Fig. 3 with references to the relevant equations. The operator $\mathcal{D}_{\downarrow 2}$ indicates downsampling by a factor of 2, with $\mathcal{G}_\sigma^* \mathbf{X}$ indicating convolution of the data with a low pass filter, namely a Gaussian with $\sigma = 0.5$ and kernel size of 3 pixels. The counters n and L represent, respectively, the number of datasets and number of levels.

Initialization requires that the number of classes $\mathcal{K}$ be either estimated using, for example, a minimum description length (Leclerc, 1989) type method or chosen heuristically. Currently it is manually specified (typically 3 or 4 classes). There is no requirement that all datasets have the same number of classes. Otherwise the algorithm is unsupervised.

The choice of the number of scales is difficult. For the purposes of registration a large number of scales is desirable to avoid convergence to local minima and allow a single (scaled) pixel translation to move further. However, for the purposes of segmentation we desire fewer as iterative isotropic blurring of an image weakens boundaries and averages regions. This can lead to errors in segmenta-

set $\mathcal{K} = manual\ initialisation.$
for l = 1 **to** n
  for l = 1 **to** L $\quad \mathbf{X}_i^l = \mathcal{D}_{\downarrow 2}\left[\mathcal{G}_\sigma * \mathbf{X}_i^{l-1}\right]$ $\quad$ **end**
  for $x \in \mathbf{X}_n^L \quad x = \text{random}(k \in \mathcal{K}_n)$ $\quad$ **end**
  (Initialize μ_k) $\quad$ **for** $k \in \mathcal{K}_n \quad \mu_k = \sum_{x \in \mathbf{X}_n^L} x.\delta(s=k) / \sum_{x \in \mathbf{X}_n^L} \delta(s=k)$ $\quad$ **end**
  (Perform $\mathcal{K}$-means) $\quad \arg^{min}_{\mu_k, k \in \mathcal{K}} \left(\sum_{x \in \mathbf{X}_n^L} arg^{min}_k \parallel x - \mu_k \parallel \right)$
  $\forall k$ Calculate ω_k, σ_k given μ_k.
end
for l = L **to** 0
  set $\beta_{field} = \frac{\beta^0_{field}}{1+l}, \beta_{parent} = \frac{\beta^0_{parent}}{2(1+l)}$
  do (loop until converged at level)
  if(rigid) **do**
    if(separate) **foreach** $i \in$n calculate $S_i = \arg^{max}_{s \in S_i} \ln \mathcal{P}(S_i/\mathbf{X}_i)$ (equation 2)
    else (combined) calculate $\mathcal{S} = \arg^{max}_{s \in \mathcal{S}} \ln \mathcal{P}(\mathcal{S}/\mathbf{X}_1, \mathbf{X}_2^{T_2}, \cdots, \mathbf{X}_n^{T_n})$ (equation 1
    end
    (Perform Registration) $\mathcal{E} = \arg^{min}_{\mathcal{T}}(S_1, S_2^T)$, Powell method (equation 11)
  while$(\mathcal{E} < \mathcal{E}_{previous})$
  else (nonrigid) calculate $\arg^{max}_{s,\mathbf{t}} \mathcal{P}(\mathcal{S}, \mathcal{T})$ using ICM. (equation 14)
  end
  foreach $k \in \mathcal{K}$ Update θ_k given $\mathcal{P}(\mathcal{S}_i^l/\mathbf{X}_1, \mathbf{X}_2^{T_2}, \cdots, \mathbf{X}_n^{T_n})$.
  while(change above threshold) (see equation 15)
  if(l>0) (Initialize Next Level)
    $(T_x^{l-1} = 2T_x^l, T_y^{l-1} = 2T_y^l, T_\theta^{l-1} = T_\theta^l)$
    foreach $i \in$n $\mathcal{S}_i^{l-1} = \mathcal{S}_i^l$.
    foreach $k \in \mathcal{K}$ Update θ_k assuming $\mathcal{P}(\mathcal{S}_i^{l-1}) = \mathcal{P}(\mathcal{S}_i^l/\mathbf{X}_1, \mathbf{X}_2^{T_2}, \cdots, \mathbf{X}_n^{T_n})$.
  end
end

Fig. 3. The complete combined segmentation and registration algorithm.

tion at a low resolution being propagated down the pyramid. The number of scales has been set at three including the original image. This prevents too much drift in the distribution means. The pyramid may also be constructed using median filtering, which can perform better for particularly noisy images.

An additional Markov prior, with the same functional form as the spatial prior of Eq. (6), is used to represent the parent–child relationship in the hierarchy. β therefore has two different values, the parent β_{parent} being half that of the spatial β_{field}. As the image is downsampled by a factor of two in each direction between levels, each parent has four children except at the edges. Given that the data are not of infinite extent, points inevitably exist where multiple data are not available for all points. Where this case exists, the single GMM is used and β is reduced by a factor of $\sqrt{2}$. The non-rigid transform regularization parameter $\sigma_t = 2$.

In (Maes et al., 1997) the use of partial volume (bilinear) interpolation is advocated over nearest neighbour interpolation for constructing the joint histogram for registration. The proposed algorithm agrees in using bilinear interpolation to construct the joint class histogram but uses nearest neighbour interpolation for interpolating data for the segmentation likelihoods in Eq. (8) (Greig et al., 1989). As Fig. 4 shows, bilinear interpolation introduces an offset to the mean and variance. Although this is beneficial within a region, as it reduces the variance similarly to low pass filtering, it is not so at the edges. Nearest neighbour interpolation is better for ensuring that incorrect data estimates are not introduced to the segmentation.

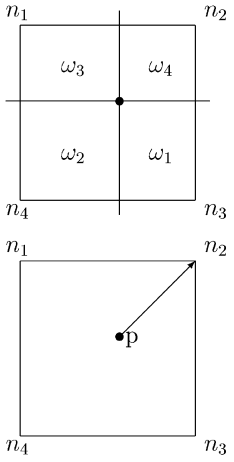


Fig. 4. Nearest neighbour versus bilinear interpolation. The expected mean and variance across a single pixel with different neighbouring distributions.

The algorithm terminates on one of two conditions. If the segmentation model parameter set θ is being updated the algorithm terminates when θ converges to a user defined tolerance; $\delta_{\text{change}} < 0.05$. This change is assessed as the weighted change in distribution parameters at the nth iteration:

$$\delta_{\text{change}} = \sum_k \omega_k \left(\frac{\mu_{k,n+1} - \mu_{k,n}}{\sqrt{\sigma_{k,n+1}\sigma_{k,n}}} \right). \quad (15)$$

In addition, the algorithm terminates when there is no change in the registration parameters as this implies that the segmentation has converged.

Table 1
Algorithm robustness with respect to β^0_{parent} ($\beta^0_{\text{field}} = 1$, $\sigma_{\text{noise}} = 60$, $K = 4$)

	β^0_{parent}	0	0.5	1	1.5	5	10
1D HMRF	FCC	0.88	0.90	0.92	0.92	0.94	0.93
	LSE μ	0.43	0.45	0.46	0.48	0.60	0.71
	LSE σ	0.60	0.60	0.60	0.63	0.73	0.89
	HD	3.68	3.52	3.67	3.86	4.81	5.79
JS&R	FCC	0.93	0.93	0.93	0.94	0.94	0.93
	LSE μ	0.35	0.36	0.37	0.38	0.48	0.65
	LSE σ	0.52	0.53	0.54	0.54	0.61	0.76
	HD	3.19	3.41	3.41	3.43	3.77	5.25

5. Results

Results are presented on simulated Gaussian datasets to illustrate the principle of our method, as well as some of the issues in JS&R. Further examples are given on MR brain images and a sequence of mouse cardiac MR images.

5.1. Synthetic images

Fig. 5(a,d) shows two synthetic images. These have been created using Gaussian classes with means, for (a) of [50, 100, 150, 200] (gray levels) and for (b) [200, 150, 100, 50] (gray levels). The classes have standard deviation of ~45 (gray levels). Note that the second image simulates a 'multi-modality' version of the first. These images are examples of those used in testing the robustness and accuracy of the algorithm on data with known ground truth.

Misregistrations are simulated by giving the JS&R algorithm different initial translations and rotations. An example obtained for correct recovery of parameters $\theta = 10°$, $T_x = 8$, $T_y = 13$ is shown in Fig. 5. Varying quantities of noise are added to the images to simulate different quality data. The JS&R algorithm is compared to three standard algorithms. To examine segmentation accuracy, comparison is made to an HMRF using a GMM for single intensity data. An example of segmentation is shown in Fig. 5(b,e) for the sample images. A further comparison is made with an HMRF using the multimodality data and a correct registration. This provides the bounding accuracy to which it is hoped JS&R will be close. In total, ~1500 image pairs have been processed for varying initial transforms. This equates to ~30 for each initial position at each of the four noise levels. The measures used to evaluate the error are defined in Appendix C.

Tables 1 and 2 show the variability of the segmentation

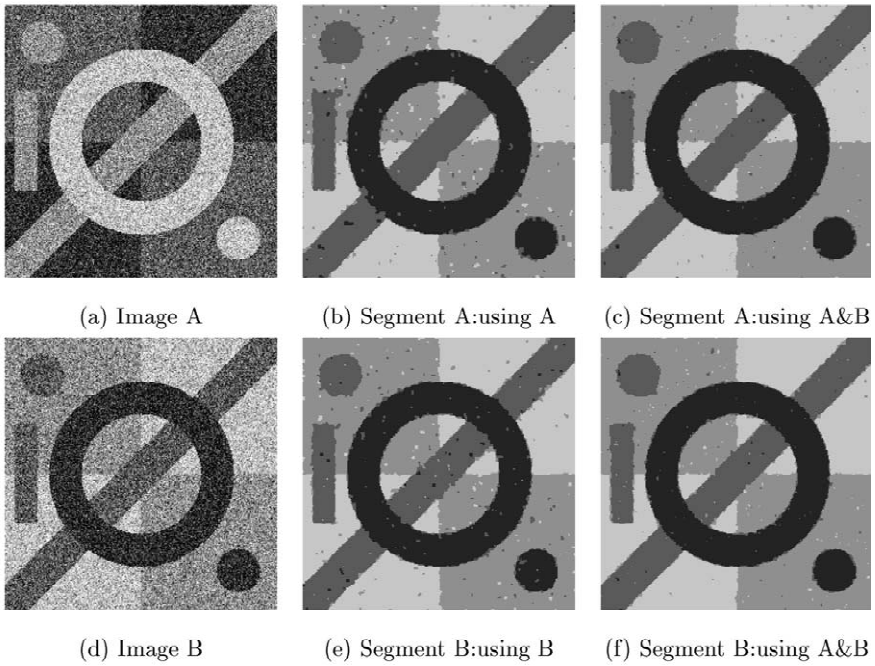


Fig. 5. 2 images with 4 classes, Gaussian noise (size 250×250 pixels).

Table 2
Algorithm robustness with respect to β^0_{field} ($\beta^0_{\text{parent}} = 0.5$, $\sigma_{\text{noise}} = 60$, $K = 4$)

Algorithm	β^0_{field}	0	0.5	1	1.5	5	10
1D HMRF	FCC	0.71	0.81	0.90	0.93	0.96	0.94
	LSE μ	0.30	0.38	0.45	0.49	0.74	1.11
	LSE σ	0.49	0.53	0.60	0.65	0.98	1.84
	HD	2.32	2.94	3.52	3.98	7.56	13.1
JS&R	FCC	0.81	0.88	0.93	0.95	0.96	0.93
	LSE μ	0.28	0.33	0.36	0.39	0.72	1.21
	LSE σ	0.46	0.48	0.53	0.55	1.00	1.93
	HD	2.26	2.72	3.41	3.54	8.03	12.5

accuracy with the two key parameters β_{parent}, β_{field}. As would be expected, the parental prior is more robust than the spatial. It is important to note that although segmentation accuracy increases through the table, it is mirrored by an increase in delocalization of boundaries. Errors within regions are being removed, but at boundaries they are being enhanced. JS&R appears less affected by this in the parameter region up to $\beta \sim 1.5$. Setting β chooses the point where evidence from n neighbours outweighs the pixel likelihood. A high value of β is undesirable as it tends to oversmooth fine detail at boundaries.

The registration is tested against a mutual information algorithm using image intensity. As with the segmentation HMRFs, all parameters have been kept identical where they exist to be set. For varying initial conditions, the fraction of correct recoveries of the transform is assessed. Correct is defined as $\theta \pm 0.1°$, $|T_x + T_y| < 0.2$ pixels. Table 3 shows the comparison between intensity MI and the JS&R joint class entropy criterion. As can be seen, at low levels of noise there is little difference between the two, but at higher noise class-based MI is more reliable. As noise increases further, $\sigma \geqslant 75$, both start to fail more significantly. Class-based MI remains more robust. In summary, for the entire set of registrations tested 78.4% were recovered correctly by the intensity MI algorithm and 96% by the combined algorithm.

Table 4 compares the recovered segmentations for the JS&R algorithm to a 1D HMRF and a 2D MRF given the known registration. As can be seen, the JS&R algorithm achieves a rate of accuracy which is higher than the 1D HMRF and is similar to the 2D HMRF.

5.2. Algorithm complexity

Although it was expected that combining segmentation and registration would increase the computational complexity this has not proved to be the case. A breakdown of timings for images of increasing size, 100^2, 200^2 and 500^2 for 3 and 255 classes/gray levels is shown in Table 5. The tests were performed on a Pentium 3 700 MHz with 512 MB ram. As can be seen from Table 5 a significant saving is made in calculating the entropy with fewer classes. This

Table 3
Fraction of recovered registrations class vs. intensity mutual information ($\beta^0_{\text{field}} = 1$, $\beta^0_{\text{parent}} = 0.5$, $\sigma_{\text{noise}} = 60$, $K = 4$)

θ	σ $\lvert T_x + T_y \rvert$	JS&R				Intensity MI			
		15	30	45	60	15	30	45	60
5	0	1.00	1.00	1.00	1.00	1.00	1.00	0.81	0.66
	5	1.00	1.00	1.00	1.00	0.98	0.86	0.81	0.68
	10	1.00	1.00	1.00	1.00	0.98	0.72	0.63	0.54
	15	0.96	0.96	0.96	0.96	1.00	0.79	0.72	0.57
10	0	1.00	1.00	1.00	0.98	0.97	1.00	0.81	0.63
	5	0.98	1.00	1.00	0.94	0.98	0.94	0.78	0.70
	10	1.00	1.00	1.00	0.92	1.00	0.73	0.60	0.52
	15	0.98	1.00	0.80	0.80	1.00	0.85	0.62	0.57
15	0	1.00	1.00	0.84	0.84	0.97	0.86	0.59	0.53
	5	0.98	0.96	0.88	0.92	0.98	0.86	0.73	0.65
	10	1.00	0.80	0.72	0.68	0.98	0.66	0.52	0.52
	15	0.98	0.92	0.88	0.72	0.98	0.79	0.60	0.55

Table 4
JS&R segmentation accuracy vs. 1D HMRF and 2D MRF given correct registration ($\beta^0_{\text{parent}} = 0.5$, $\beta^0_{\text{field}} = 1$, $K = 4$)

σ_n	1D HMRF			JS&R			2D HMRF (known transform)		
	FCC	μ_{LSE}	σ_{LSE}	FCC	μ_{LSE}	σ_{LSE}	FCC	μ_{LSE}	σ_{LSE}
15	0.99	0.02	0.14	0.99	0.04	0.20	0.99	0.02	0.15
30	0.98	0.14	0.34	0.98	0.13	0.34	0.99	0.12	0.26
45	0.94	0.29	0.49	0.96	0.23	0.43	0.96	0.22	0.40
60	0.89	0.46	0.59	0.93	0.36	0.53	0.94	0.30	0.48

Table 5
Timings for combined segmentation and rigid registration algorithm computational cost breakdown (times given in milliseconds per iteration)

Image size	100×100		200×200		500×500	
No. classes/grayscales	3	255	3	255	3	255
Calculation of mappings	2	2	10	10	44	44
Form histogram (PV interpolation)	7	7	43	43	267	267
Form histogram (NN interpolation)	2	2	10	10	48	48
Calculation of entropy	0.2	5.5	0.2	5.5	0.2	5.5
Total time per iteration	9.2	14.5	53.2	58.5	311.2	316.5

results from fewer log calculations. As the MRF typically takes between 0.15 and 1.5 seconds to run depending upon image size and the number of iterations for ICM convergence (typically 4–7) a time saving is normally realized. As the proposed algorithm uses a multiresolution search strategy, fewer iterations will be necessary at high resolution. This results from convergence having been primarily obtained at low resolution. For small datasets a factor of 10 is often achievable.

It should also be noted that one reason MI typically uses bilinear (PV) interpolation is that it provides smoother histogram construction which helps reduce the number of local minima. As the class-based criterion is more robust in this respect there is no reason that nearest-neighbour interpolation could not be used instead, which would increase the time saving.

5.3. Phantom images

In Fig. 6 the results of applying rigid registration to MR phantom brain images taken from brainweb (http://www.bic.mni.mcgill.ca/) are shown. T1 and T2 weighted images with slice thickness 1 mm, noise of 9% and intensity inhomogeneity of 20% are shown. It is difficult to validate this example as the images are generated from a fuzzy model of the brain and so binarization does not yield an unambiguous class map making ground truth hard to obtain. However, it can be seen visually that the JS&R segmentations maintain better separation of the sulci (i.e. between the white and gray matter) and retain finer detail than the separate segmentations. This detail is lost by the separate algorithm owing to insufficient separation in the 1D likelihood model. The simulated misregistration was recovered by both the separate registration process and JS&R algorithm. For ease of visual comparison a one-to-one match has been enforced.

5.4. Medical images

Fig. 7 shows five frames covering approximately one quarter of a cardiac cycle, of a mouse heart MR sequence. The sequence was obtained from an 11.7 Tesla machine.

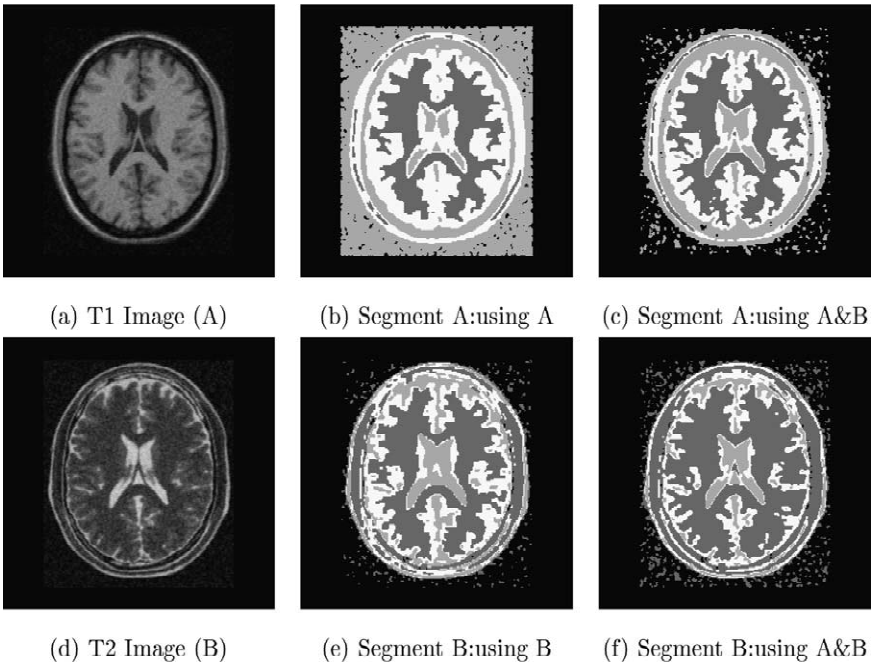


Fig. 6. Segmentation and registration of multimodal brain images; separate versus combined results.

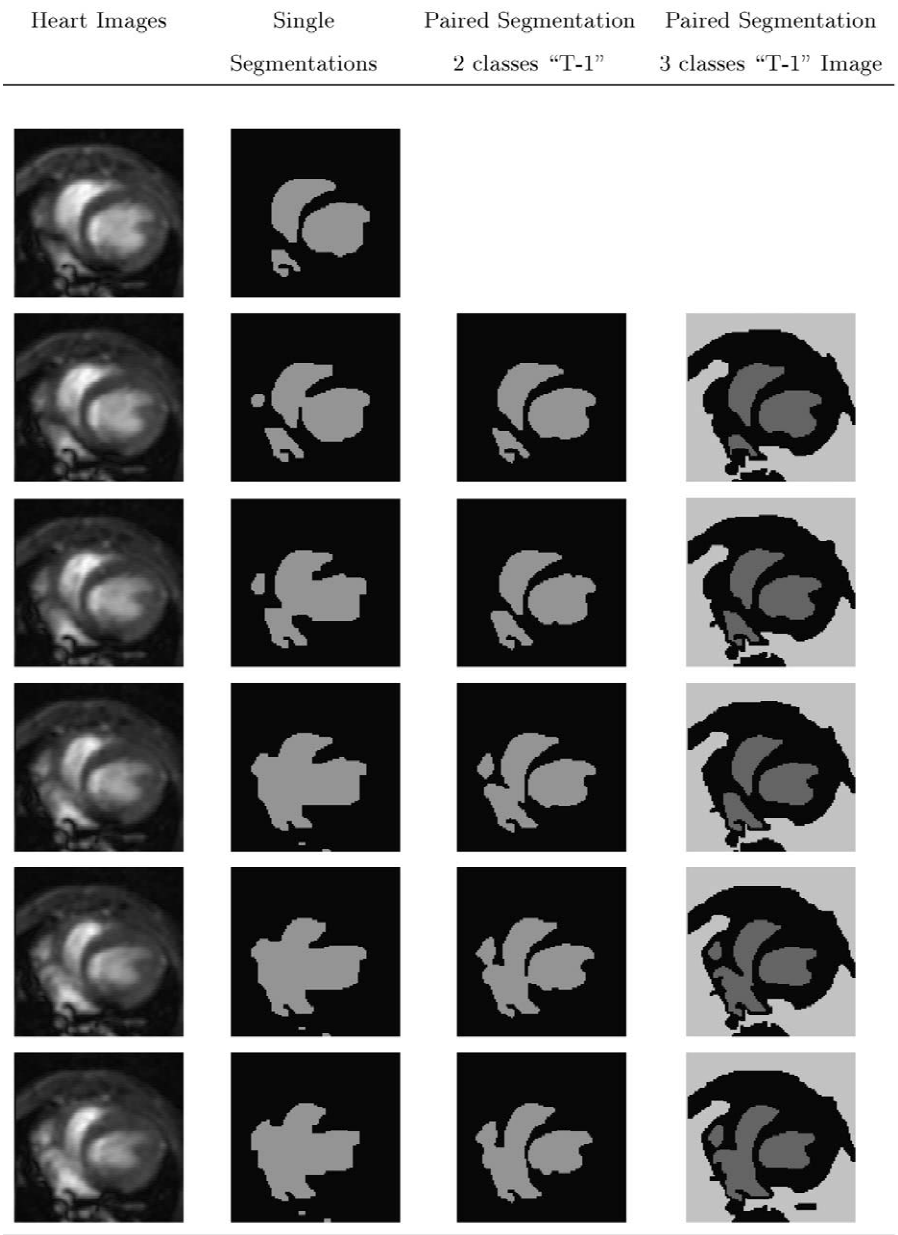


Fig. 7. Segmentations, of ~1/4 cycle, mouse heart MRI sequence (size: 100×100 pixels).

Each voxel is ~0.1×0.1×1.0 mm^3 in size with temporal resolution of ~5 ms. The first column shows the MRI images, the second the single dataset segmentations. The third and fourth columns show the paired segmentations obtained using JS&R (non-rigid) on pairs of adjacent images with two and three classes, respectively. The combined segmentations appear better visually than the separate segmentations; maintaining separation of left and right ventricles. Segmentation accuracy was compared to a manual segmentation of the ventricles. Assessing the error in non-rigid registration is difficult as ground truth is difficult to establish. For the registration, one possibility is that, if the noise field is Gaussian, subtraction of the registered image projected into the frame of the source image, from the source image should leave a random field of white noise behind. To avoid bias, the means of the images are shifted to zero and they are normalized before subtraction. The registrations obtained from conventional mutual information and the combined JS&R algorithms produced images with very similar appearance. Subtracting the registered images from the source left no statistically significant difference. As Table 6 shows, whilst there is only a small increase in the fraction of correctly classified pixels the error and variance in the location of the boundary are substantially reduced, as is the Hausdorff distance. The improvement in correctly classified pixels is small as interpolation near borders can give misleading information. The Hausdorff distance reduces principally as

Table 6
Error in 2 class segmentations of mouse heart sequence (averages over 22 frames)

Error measure	Separate Mean (±S.D.)	JS&R Mean (S.D.)
Fraction correctly classified	0.9596 (0.015)	0.9638 (0.004)
LSE in boundary (pixels)	1.5864 (0.577)	0.9961 (0.299)
Variance in boundary (pixels sqr)	2.2827 (1.111)	1.0115 (0.533)
Hausdorff distance (pixels)	10.9298 (5.608)	5.2957 (2.963)

separation between left and right ventricles is maintained by the combined algorithm and not always by the separate.

6. Discussion and summary

The aim of this paper has been to demonstrate the advantages in applying a combined segmentation and registration method, developing foundations for such an algorithm and showing potential applications. Several key aspects which can be improved have also been identified.

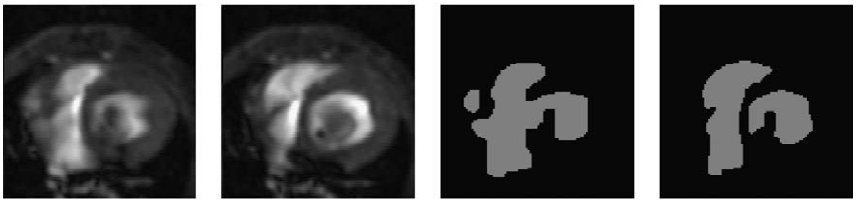
The proposed algorithm performs well for two reasons. Firstly, the Markov prior introduces spatial regularization into the registration through the class-based entropy measure of Eq. (11). This has the additional benefit of reducing the computational load of registration which offsets the additional cost of segmentation. Secondly, the minimization of entropy of classes leads to good matching where the additional information improves segmentation accuracy. In addition, the JS&R algorithm retains the information theoretic criterion advantage of being modality independent.

The results of Figs. 5–7 illustrate an improvement from the combination of information. Essentially this combination makes an additional set of features available. In our algorithm these feature sets have been limited solely to intensity. In a single dataset a similar effect can be achieved by computing wavelet coefficients or co-occurrence statistics. The advantage of the feature(s) from the second data set versus additional feature(s) computed on the same dataset is that as the noise is motion-decorrelated between datasets it should be more robust and reliable. The Gaussian model currently employed for segmentation is known to fail on texture segmentation and exhibit poor performance on various imaging modalities, e.g. ultrasound. Developing models to make use of scale-based features from both images might help overcome this limitation.

Possibly the most important, though difficult to analyse, issues are convergence and robustness. Theoretically, it can be shown that the expected classification error resulting from a GMM-MRF with four neighbour cliques is greater than or equal to ~40% of that of the maximum likelihood (ML) error for the same GMM without the spatial prior (Appendix B). It would therefore seem reasonable to expect a corresponding improvement in robustness to noise for the MAP registration criteria over the ML. In practice, the improvement may not attain the theoretical owing to interpolation errors, the GMM's failings in modelling real world problems and the theoretical requirement that β be allowed to approach infinity. However, it has been demonstrated that practically significant improvement in accuracy and robustness is realized.

The greatest weakness in the method is in setting the priors β_{field}, β_{parent} and σ_t. Particularly in the case of β_{field}, too high a value leads to over smoothing of boundaries. Although the joint likelihood model $\mathcal{P}(X_1, X_2/S, T)$ is beneficial, it is not always sufficient. Two potential additions can help here. The first is construction of the multiresolution hierarchy with an anisotropic or region preserving filter, i.e. median filter. The second is to alter the prior weight according to a locally adaptive measure, such as suggested by Boukerroui et al. (2003).

Fig. 8 shows the segmentation of frame 14 of the mouse heart sequence (using frame 15 for JS&R) and one instance where the combined algorithm performs poorly. Although in Fig. 8(d) the JS&R algorithm retains separation of the left and right ventricles, unlike the separate image segmentation Fig. 8(c), the centre of the bloodpool is misclassified. This error results from the segmentation model simplicity: there is no higher level information available to model the difference in appearance and the local property of the MRF is unable to maintain region continuity. In order to be able to correctly match the images there needs to be overlap between the features locally. This is one reason for implementing a multiresolution search strategy. The effective increase in search area provided by the lower resolutions aids convergence at the finer scales. However, in Fig. 8 the intensity and class-based non-rigid criteria used are unable to cope with this fading; indeed, any region-based method would struggle.



(a) Frame 14 (b) Frame 15 (c) Separate (d) JS&R

Fig. 8. Erroneous classification resulting from fading of bloodpool.

A final consideration concerns the extension to 3D applications. The principal difference is a necessary redefinition of the Markov prior model. Depending upon the scanning method used, i.e rotating 3D probe or planar slices, the Markov field will cease to be isotropic and homogeneous as a result of the non-uniformity of resolution. Therefore, it would be necessary to determine appropriate clique functions. In the case of sparse data there would be less of an advantage in joint segmentation as improvements in accuracy stem from monomodal/multimodal compounding of information. The exact utility would depend upon the extent to which data smoothness could be assumed and compounding exists.

To summarize, results indicate that the addition of a spatial prior leads to substantially greater robustness in rigid registration and the combination of data improves segmentation accuracy. There is also an increase in speed for the case of combined segmentation and rigid registration. This suggests the combination of segmentation and registration is worth pursuing in future.

7. Further references

For further reading see (Bansal et al., 1998; Brailean and Katsaggelos, 1995a, 1995b; Brailean et al., 1995; Papademetris et al., 2001; Roche et al., 2000).

Acknowledgements

PPW gratefully acknowledges the financial support of the UK EPSRC for funding this research and Juergen Schneider for providing the MR data. This paper is an expanded version of a conference article presented at MICCAI 2002.

Appendix A. Improved separability of functions in higher dimensions

If we can prove, for any N-dimensional function $\mathscr{F}$, that the error of maximum likelihood (ML) classification is upper-bounded by the 1D case, then this proves that ML classification from the fusion of registered datasets yields equal or more accurate results than those possible for any single dataset.

Consider 1D functions $\mathscr{F}_1(x_1)$, $\mathscr{F}_2(x_1)$, with ML error $\delta\mathscr{E}_1$ (about point a)

$$\delta\mathscr{E}_1 = \int_{a-\delta a}^{a+\delta a} \min\left[\mathscr{F}_1(x_1), \mathscr{F}_2(x_1)\right] \mathrm{d}x_1. \tag{A.1}$$

In 2D these form, after fusion with other data, functions $\mathscr{G}_{2,1}(x_1, x_2)$, $\mathscr{G}_{2,2}(x_1, x_2)$. The fractional error for the same data becomes

$$\delta\mathscr{E}_2 = \int_{a-\delta a}^{a+\delta a}\int_{-\infty}^{\infty} \min\left[\mathscr{G}_{2,1}(x_1, x_2), \mathscr{G}_{2,2}(x_1, x_2)\right] \mathrm{d}x_2\, \mathrm{d}x_1. \tag{A.2}$$

Iff the data are independent, then $\mathscr{G}_{2,(1\cup 2)}(x_1, x_2) = \mathscr{F}_{(1\cup 2)}(x_1)\mathscr{G}(x_2)$ and Eq. (A.2) becomes

$$\begin{aligned}\delta\mathscr{E}_2 &= \int_{a-\delta a}^{a+\delta a}\int_{-\infty}^{\infty} \min\left[\mathscr{G}(x_2)\mathscr{F}_1(x_1), \mathscr{G}(x_2)\mathscr{F}_2(x_1)\right] \mathrm{d}x_2\, \mathrm{d}x_1 \\ &= \int_{-\infty}^{\infty} \mathscr{G}(x_2)\, \mathrm{d}x_2 \int_{a-\delta a}^{a+\delta a} \min\left[\mathscr{F}_1(x_1), \mathscr{F}_2(x_1)\right] \mathrm{d}x_1 \\ &= \int_{a-\delta a}^{a+\delta a} \min\left[\mathscr{F}_1(x_1), \mathscr{F}_2(x_1)\right] \mathrm{d}x_1. \end{aligned} \tag{A.3}$$

Hence, the error for $2\mathscr{D}$ is equal to that for $1\mathscr{D}$. This also implies that the ratio $\mathscr{F}_1(x_1)/\mathscr{F}_2(x_1) = \mathscr{K}$, a constant. However, if $\mathscr{G}_{2,(1\cup 2)}(x_1, x_2) \neq \mathscr{F}_{(1\cup 2)}(x_1)\mathscr{G}(x_2)$ then the ratio $\mathscr{G}_{2,1}(x_1, x_2)/\mathscr{G}_{2,2}(x_1, x_2) \neq \mathscr{K}$. The maximum ML misclassification must be $\delta\mathscr{E}_1$, as the maximum union of $\mathscr{F}_1(x_1)$, $\mathscr{F}_2(x_1)$ is min $\left[\mathscr{F}_1(x_1), \mathscr{F}_2(x_1)\right]$. Consequently, extending this logic to the whole domain:

$$\text{Iff } \frac{\mathscr{G}_{\mathscr{N},1}(x_1, x_2, \ldots, x_{\mathscr{N}-1}, x_{\mathscr{N}})}{\mathscr{G}_{\mathscr{N},2}(x_1, x_2, \ldots, x_{\mathscr{N}-1}, x_{\mathscr{N}})} \geqslant \cdots \geqslant \frac{\mathscr{G}_{2,1}(x_1, x_2)}{\mathscr{G}_{2,2}(x_1, x_2)} \geqslant \frac{\mathscr{F}_1(x_1)}{\mathscr{F}_2(x_1)} \quad \text{then } \delta\mathscr{E}_2 \leqslant \delta\mathscr{E}_1.$$

Appendix B. Expected error reduction of MRF with respect to the ML error

Consider a Gaussian model of parameter $\mathbf{x}$ with mean μ_i, variance σ_i and weight ω_i;

$$y_i = \frac{\omega_i}{(\sqrt{2\pi}\sigma_i)} \mathrm{e}^{-1/2(\mathbf{x}-\mu_i/\sigma_i)^2}.$$

Models y_1 and y_2 will intercept at $\mathbf{x}_a$, $\mathbf{x}_b$, obtained from the solution to

$$(\sigma_2^2 - \sigma_1^2)\mathbf{x}^2 + 2(\mu_2\sigma_1^2 - \mu_1\sigma_2^2)\mathbf{x} + \left[\mu_1^2\sigma_1^2 - \mu_2^2\sigma_1^2 - 2\ln\left(\frac{\omega_1\sigma_2}{\omega_2\sigma_1}\right)\right] = 0. \tag{B.1}$$

The maximum likelihood classification error is simply

$$\mathscr{E} = \int_{-\infty}^{\min(x_a,x_b)} \min(y_1, y_2)\,dy + \int_{\min(x_a,x_b)}^{\max(x_a,x_b)} \min(y_1, y_2)\,dy + \int_{\max(x_a,x_b)}^{\infty} \min(y_1, y_2)\,dy.$$

The effect of an MRF is essentially that locally it alters the relationship between y_i by using the current local neighbour class to alter the weight ω_i. This shifts the intercepts in Eq. (B.1) as the weight becomes $\omega_i e^{-\Sigma_{c\in\mathscr{C}}\ \beta\mathscr{V}_c(\mathscr{S})}$. As $\beta\to\infty$ the intercept for all cliques apart from the single exception where the prior clique energies $\Sigma_{c\in\mathscr{C}}\ \mathscr{V}_c(\mathscr{S})$ are equal for both y_1, y_2 will tend to $\pm\infty$. Hence the errors for those cliques will be zero. For the clique with equal energies, the error will be the same as for the MLE, excepting that this clique is now *less probable*. Its expected frequency of occurrence is obtained from the binomial distribution as, where $\mathscr{B}_n^N = N!/n!(N-n)!$ is the binomial coefficient, $\mathscr{B}_{N/2}^N / \Sigma_0^N\ \mathscr{B}_n^N$. Therefore the MRF error can be estimated as approximately this fraction of the MLE. This assumes that the lattice is homogeneous.

For the special case of a four-neighbourhood, this is $6/16 \sim 0.375$.

Appendix C. Definitions of error measures used

The error measures used are defined as follows.

The *Hausdorff measure* (HD) calculates the maximum of the minimum distances between two curves or sets of numbers. If two sets of points correspond to two curves $\mathscr{P} = \{p_1, p_2, \ldots, p_n\}$ and $\mathscr{Q} = \{q_1, q_2, \ldots, q_n\}$, then the Hausdorff Error $\mathscr{E}_{\mathscr{H}}$ is

$$\mathscr{E}_{\mathscr{H}} = \mathrm{Arg}_i^{\mathrm{Max}}\left[\mathrm{Arg}_j^{\mathrm{Min}}\|\mathscr{P}_i - \mathscr{Q}_j\|\right]. \tag{C.1}$$

The LSE is the average of the distance between two curves, which in the case of segmentation are the boundary between regions for the ground truth and test segmentation:

$$\mathscr{E}_{\mathrm{LSE}} = \frac{1}{N}\sum_{i=1}^{N}\left[\mathrm{Arg}_j^{\mathrm{Min}}\|\mathscr{P}_i - \mathscr{Q}_j\|\right]. \tag{C.2}$$

Fraction Correctly Classified (*FCC*): if the ground truth for pixels $s\in\mathscr{S}$ is known and it is compared to a trial segmentation $r\in\mathscr{S}$, then the fraction of correctly classified pixels is

$$FCC = \frac{1}{\mathscr{S}}\sum_{\mathscr{S}}\delta(s,r) \quad \text{where } \delta(s,r) = \begin{cases} 0: & s\neq r, \\ 1: & s = r. \end{cases} \tag{22}$$

References

Bansal, R., Staib, L.H. et al., 1998. A novel approach for the registration of 2D portal and 3D CT images for treatment setup verification in radiotherapy. In: MICCAI, pp. 1075–1086.

Bansal, R., Staib, L.H. et al., 1999. Entropy-based multiple portal to 3D CT registration for prostate radiotherapy using iteratively estimated segmentation. In: MICCAI, pp. 567–578.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B 36, 192–236.

Besag, J., 1986. On the statistical analysis of dirty pictures. J. R. Stat. Soc. Ser. B 48 (3), 259–302.

Bishop, C., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Boukerroui, D., Baskurt, A. et al., 2003. Segmentation of ultrasound images—multiresolution 2D and 3D algorithm based on global and local statistics. Pattern Recogn. Lett. 24 (4), 779–790.

Brailean, J.C., Katsaggelos, A.K., 1995a. Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences. IEEE Trans. Image Process. 4 (9), 1236–1251.

Brailean, J.C., Katsaggelos, A.K., 1995b. A recursive nonstationary MAP displacement vector field estimation algorithm. IEEE Trans. Image Process. 4 (4), 416–429.

Brailean, J.C., Kleihorst, R.P. et al., 1995. Noise reduction filters for dynamic image sequences: a review. Proc. IEEE 83 (9), 1272–1290.

Brown, L.G., 1992. A survey of image registration techniques. ACM Comput. Surveys 24 (4), 325–376.

Clark, M., Hall, L. et al., 1998. Automatic tumor segmentation using knowledge-based techniques. IEEE Trans. Med. Imag. 1 (2), 187–201.

Cootes, T.F., Taylor, C.J., 2001. Statistical models of appearance for medical image analysis and computer vision. Proc. SPIE Med. Imag., Vol. 4322, pp. 236–248.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE PAMI 6 (6), 721–741.

Greig, D.M., Porteos, B.T., Seheult, A.H., 1989. Exact maximum a posteriori estimation for binary images. J. R. Stat. Soc. Ser. B 51 (2), 271–279.

Hurn, M.A., Mardia, K.V. et al., 1996. Bayesian fused classification of medical images. IEEE Trans. Med. Imag. 15 (6), 850–858.

Leclerc, Y.G., 1989. Constructing simple stable descriptions for image partitioning. IJCV 3, 73–102.

Maes, F., Collignon, A. et al., 1997. Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imag. 16 (2), 187–198.

Maintz, J.B.A., Viergever, M.A., 1998. A survey of medical image registration methods. Med. Image Anal. 2 (1), 1–36.

McLaughlin, R.A., Hipwell, J. et al., 2002. A comparison of intensity-based registration and feature-based registration for neurointerventions. In: Proceedings MICCAI, Japan, pp. 517–524.

Papademetris, X., Onat, E.T. et al., 2001. The active elastic model. Inform. Process. Med. Imag, June.

Pennec, X., 1998. Towards a generic framework for recognition based on uncertain geometric features. J. Comput. Vis. Res. 1 (2), 58–87.

Pizer, S., Fritsch, D. et al., 1999. Segmentation, registration and measurement of shape variation via image object shape. IEEE Trans. Med. Imag. 18 (10), 851–865.

Press, W.H., Teukolsky, S.A. et al., 1992. Numerical Recipes in C. Cambridge University Press.

Roche, A., Malandain, G., Ayache, N., 1999. Unifying maximum likelihood approaches in medical image registration. In: INRIA, France, Tech Report RR-3741.

Roche, A., Pennec, X., Malandain, G., Ayache, N., Ourselin, S., 2000. Generalized correlation ratio for rigid registration of 3D ultrasound with MR images. In: INRIA, France, Tech. Report RR-3980.

Wells, W., Viola, P. et al., 1996. Multi-modal volume registration by maximization of mutual information. Med. Image Anal. 1 (1), 35–51.

Yezzi, A., Zollei, L., Kapur, T., 2001. A variational framework for joint segmentation and registration. In: IEEE Proc. MMBIA, pp. 44–52.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imag. 20 (1), 45–57.