Automated Analysis of Deformable Structure in Groups of Images

V Petrović, T Cootes, C Twining, A Mills and C Taylor
Imaging Science and Biomedical Engineering
University of Manchester
M13 9PT, UK
v.petrovic@manchester.ac.uk
http://www.isbe.man.ac.uk

Abstract

We describe an approach for automated analysis of deformable objects which extracts structure information from groups of images containing different examples of the object with a particular application to human imaging. The proposed analysis framework simultaneously segments and registers a set of images, incrementally constructing a model of the composition of the object. By fitting an appropriate intensity distribution model to the image we obtain a soft segmentation which allows us to explicitly model the construction of each pixel from constituent image segments, rather than its expected intensity. This effectively decouples the model from the effects of the imaging system and varying statistics in different examples. When estimating the optimal deformation field for each example, the original image is compared to a reconstruction, generated using the composition model and its intensity distribution parameters for each segment (i.e. an estimate of how the model would appear given the imaging conditions for that image). In the paper we describe the algorithm in detail and show results of applying it to two sets of medical images of different anatomies taken with different imaging modalities. We present quantitative results demonstrating that the proposed algorithm is more powerful than current state of the art methods at extracting structural information such as spatial correspondences across groups of images with varying statistics.

1 Introduction

This paper proposes an automated approach for analysing, understanding and representing deformable object structure in groups of images, with a particular application to medical imaging and biometrics. The human body is an abundant source of objects that share a common anatomical structure but exhibit an almost infinite number of shape and appearance variations. Generally, given a set of images of different examples of an object with a deformable structure, we would like to derive in an automated manner (without user intervention) the following:

- a dense spatial and structural correspondence between the various examples (registration)

- a consistent composition of the pixels in each example image into different constituent parts of the structure (segmentation)
- a statistical representation of the variability of shape and appearance of the structure across the set (modelling)

Furthermore, an explicit advantage would be if all of the above could be achieved in an efficient and robust (converging) manner. There has already been considerable research into techniques that aim to reach each of the listed goals independently. Non-rigid image registration, and in particular *groupwise* methods provide a method of deriving a dense, spatial correspondence across sets of images [10, 1] (for a review see [14]). Direct segmentation of medical images, into different tissues for example, has also been studied extensively with methods based on pixel intensity and more advanced deformable structures [8, 13]. Finally, Statistical Shape and Appearance Models [4], are capable of capturing and describing the appearance (shape and texture). variation of the modeled structure.

A number of other works exploit the fact that a good estimate of any one aspect of the structure, a correct segmentation, registration or a good model, can help derive more reliable estimates of the other components. For instance combined segmentation and registration with active contours was considered in [12] to register single objects. Maximum a posteriori segmentation using hidden Markov random fields and B-spline non-rigid registration was used for more general medical images [2]. Models of deformation have been constructed from correspondences estimated by non-rigid registration [7, 9], but it was also shown that it is possible to integrate modelling and registration more tightly [5].

In this paper we describe an automated approach which combines simultaneous segmentation, registration and modeling of structure in a single iterative framework to satisfy the requirements laid out above. The method starts with a training set of images and incrementally constructs a model of the composition of each pixel in the common structure, rather than its expected intensity. This decouples the model from details of the imaging process and modality and allows us to deal with datasets exhibiting significant variation in intensity. Extensive qualitative and quantitative results demonstrate that the proposed algorithm is more powerful than current state of the art methods at extracting structural information such as spatial correspondences across groups of images with varying statistics.

The method is described in detail in Section 2 while results of applying it to two sets of medical images of different anatomies taken with different imaging modalities, digital radiography (X-ray, DR) and magnetic resonance (MR), are provided in Section 3. Finally we provide a discussion on the relative merits of the presented approach.

2 Method

An overview of the proposed approach is illustrated in Figure 1 showing example images from an application of the approach to MR images of the human brain. Generally, a set of N images T_i, $i = 1...N$, (the training set) is assumed to contain a common structure that consists of M distinct components whose content is defined according to some composition model F and whose intensities obey some specific distribution model with parameters θ_i. Furthermore, for the entire set, a spatial correspondence with a reference

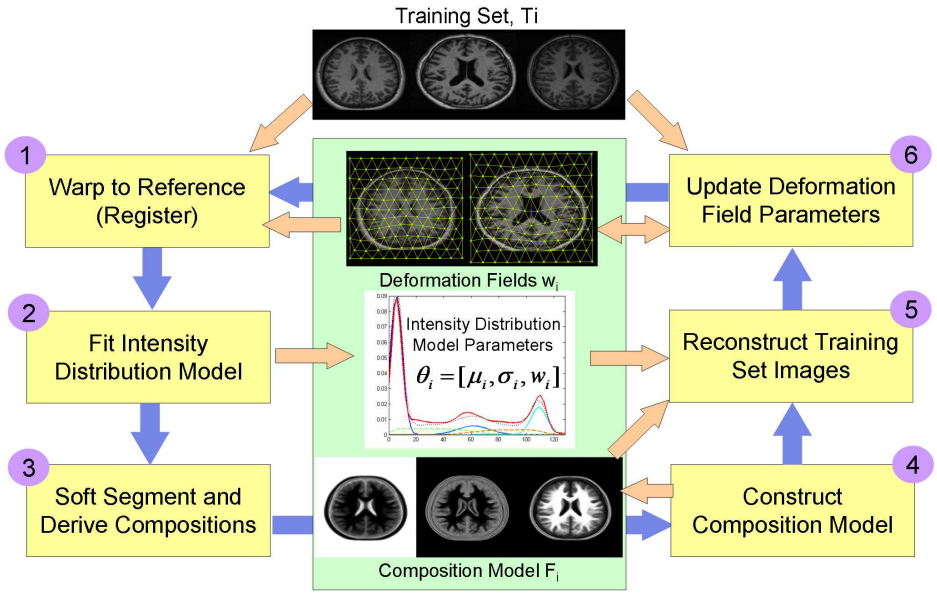

Figure 1: Outline of the proposed structure analysis algorithm: dark arrows indicate the progress of the algorithm, light arrows flow of data and the central box contains the structural information derived from the data set

(model) frame, and implicitly with each other, is assumed through a set of spatial deformation fields defined for each example in the training set, $W_i()$. Deformations are initialised as identity transformations and true correspondences $W_i()$ along with the intensity distribution model parameters θ_i and the structure composition model F are then estimated incrementally across the set in an iterative procedure as follows:

1. Warp each training image T_i into the reference frame using the current estimate of the deformation field. $T_i' = W_i^{-1}(T_i)$.
2. Fit the intensity distribution model to each image and extract parameters (means, SDs and weights) for each of the M components encoded in $\theta_i = \{\mu_{ij}, \sigma_{ij}, w_{ik}\}$), as well as distributions due to mixtures of components.
3. Use the resulting distributions to estimate the most probable composition of each pixel, and encode a set of fraction images $F_i^{(j)}$, $j = 1...M$ for each training example.
4. Combine the fraction images from all examples to construct a single composition model for the common structure, $\{\hat{F}^{(1)}...\hat{F}^{(M)}\}$.
5. Synthesize a reconstruction of each training set image S_i using the current estimates of intensity distribution parameters θ_i (μ_{ij}) and the current composition model $\hat{F}$
6. Update the current estimate of W_i to best register S_i onto T_i, minimising a suitable similarity measure, $D_{im}(T_i, W_i(S_i))$.

The stages listed above are repeated in an iterative procedure until the deformation field optimisation and the composition model converge. The reference frame defining the model shape is obtained as the mean of all individual shapes, represented through W_i. Initial identity deformation fields will contain a considerable misalignment of the examples

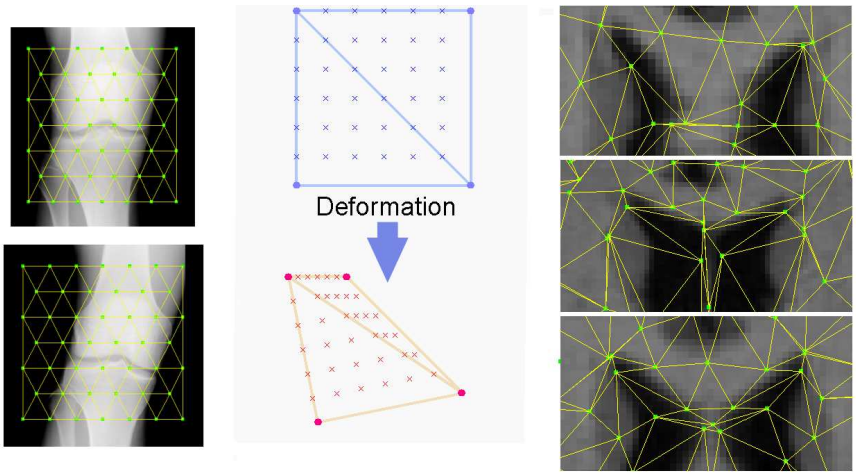

Figure 2: Piecewise affine deformation field: identity fields on two X-ray images (left), linear interpolation of a deformed shape (middle) and converged fields on corresponding areas in three MR brain images

resulting in a fuzzy composition model. However, as the algorithm progresses and correct correspondences become established both the composition model and the model shape will converge to a true, crisp representation of the underlying structure.

Note that the described process involves no construction of a shape model. Instead an explicit statistical appearance model of the structure can be constructed directly at the end of the process from the converged $W_i()$, see [5].

2.1 Establishing Correspondence

Spatial correspondence between the examples is established by defining a deformation field for each image in the training set that defines where each pixel on the reference structure is located on that image. This implicitly imposes a structural correspondence that allows equivalent locations to be found across the examples. We adopt a piece-wise affine deformation field represented as a tesselation (triangulation in 2D) of a set of control points (vertices) in space, Figure 2. Deformation is controlled by displacement of the control points, which can be both linear (e.g. affine) and highly non-linear (movements of individual points). Inside the elements the field is interpolated linearly, Figure 2, which lends efficiency and more importantly easy invertibility to this formulation at a price of limited spatial resolution and flexibility.

Deformation fields are initialised in 2D as a regular hexagonal mesh made up of equilateral triangles, see Figure 2, which provides a regular element density around each control point as opposed to a square regular mesh. The fields are then optimised in discrete stages that modify the locations of control points either in groups or individually. The details of the optimisation strategy are beyond the scope of this paper, but the general approach is to start with linear transformations (e.g. affine), followed by coarse non-rigid deformations, e.g. grid deformations [5] and progressively increase the resolution of the deformations to finish by optimising the location of each control point independently.

2.2 Segmentation

A broad segmentation of the analysed structure is achieved in two stages. First an intensity distribution model (IDM) is fit to the intensity histogram of the data in the reference frame and then a most likely composition of each pixel in each example is derived using IDM parameters. The IDM explains how the intensities in the image are related to the main components of the structure to be analysed. In principle any type of distribution model can be used within this framework but it is likely that each type of data would optimally obey a specific model. As the choice of the intensity model for a particular dataset is not central to the structure of the proposed algorithm it is not considered in detail in this paper. Instead we use relatively simple models that rely only on intensity and demonstrate the convergence power of the approach.

In general we follow [8] in assuming that each pixel in the structure is either due to one of M different components or a fractional mixture of at most two different ones. Furthermore, if we know the distributions of intensities for pure components, we can construct the distribution for a particular fractional distribution by convolution. For example, in the experiments using MR images we use a limited resolution IDM that assumes components with Normal distributions, $p_i(g) = N(g : \mu_i, \sigma_i^2)$ (consistent with white matter, grey matter and cerebro-spinal fluid/background tissue types). In this case it can be shown that the distribution for a partial volume with fraction f of tissue type i and $1-f$ of type j is given by

$$p_{ij}(g|f) = N(g : f\mu_i + (1-f)\mu_j, f\sigma_i^2 + (1-f)\sigma_j^2). \tag{1}$$

The distribution over all partial volumes containing i and j is given by

$$p_{ij}(g) = \int_{f=0}^{f=1} p_{ij}(g|f)p(f)df = \int_{f=0}^{f=1} p_{ij}(g|f)df \tag{2}$$

where we assume all values of f in the range $[0,1]$ are equally likely ($p(f) = 1$). Making the assumption that any pixel contains at most 2 different tissue types, we need only consider M pure tissue classes with distributions $p_k(g)$, $k = 1..M$, and $M(M-1)/2$ partial tissue classes (enumerated $p_k(g)$,$k = (M+1)..M_t = M(M+1)/2$). Thus the measured image intensity distribution, $h(g)$, can be approximated as a weighted sum

$$p(g : \theta) = \sum_{i=1}^{M_t} w_i p_i(g) \tag{3}$$

where $\theta = \{\mu_i, \sigma_i, w_k\}$ ($i = 1..M, k = 1..M_t$).

We thus perform an optimisation to estimate the parameters θ which optimise $D_p(p(g : \theta), h(g))$, where $D_p(p,q)$ is a suitable measure of divergence between distributions. Having estimated the probability that a pixel with intensity g belongs to class k is given by $P_k(g) = w_k p_k(g)/(\sum w_k p_k(g))$ (see Figure 3) that pixel can then be classified as belonging to class

$$k_c = \arg\max_k P_k(g). \tag{4}$$

However, we are actually interested in the estimate of the fraction of each pure class tissue (f_i,$i = 1..M$), in the pixel, not the probability of each class. If $k_c \leq M$ then the pixel is a pure tissue, so we define $f_{k_c} = 1$ and $f_{i \neq k_c} = 0$. If $k_c > M$ then the pixel is classified

as a partial volume, containing two tissues, say of type i and type j. In this case we wish to find the most likely value of the fractions for each tissue. We define

$$\begin{aligned} f_i &= \arg\max_f p_{ij}(f|g) \\ &= \arg\max_f p_{ij}(g|f)p(f)/p(g) \\ &= \arg\max_f p_{ij}(g|f) \end{aligned} \tag{5}$$

where $p_{ij}(g|f)$ is defined above in Equation 1. We then set $f_j = 1 - f_i$ and $f_{k\neq i,j} = 0$. Figure 3 shows an example of this, demonstrating that tissue probabilities are not the same as estimates of pure tissue fractions. Using this approach we compute M images, $\{F_i^{(1)}, ...F_i^{(M)}\}$, recording the fraction of each tissue type at each pixel in the normalised version of image i (that projected into the reference frame).

2.3 Composition Model Construction

The composition model defines how much of each of the components is present at any location within the structure that is being analysed. We train this model using the M fractional images from each of the N images in our set. [1] Though more detailed statistical models (eg PCA based methods) are possible, in this preliminary study we simply compute the mean of the fraction images,

$$\{\hat{F}^{(1)}...\hat{F}^{(M)}\} = \frac{1}{N}\sum_i \{F_i^{(1)}...F_i^{(M)}\}. \tag{6}$$

Further constraints could be imposed on the model, e.g. limit any pixel to have at most two non-zero fractions. Although this would directly support convergence, particularly in the early stages of the process when misalignments between different examples are still considerable we found that even the simple mean was proving powerful enough to drive the process to convergence.

2.4 Image Reconstruction

The training set is aligned by optimising a deformation field between each T_i and the model (reference frame) embodied in a reconstruction, S_i produced using the current composition model and the current estimate of the IDM parameters. Pure components exhibiting Gaussian distributions are optimally represented by their mean (μ_{ij}) while fractional pixels are represented by a sum of component means weighted by their fractions:

$$S_i = \sum_{j=1}^{M} \mu_{ij}\hat{F}^{(j)}. \tag{7}$$

For an example, see Figure 3. Essentially, S_i is an estimate of how the model would appear given the imaging conditions for T_i. Ideally S_i is a noise free version of T_i but in practice it starts blurred due to misalignments and gets progressively sharper as alignment across the set improves. Deformation parameters W_i are optimised with respect to an objective function measuring similarity between T_i and S_i in the training image frame - $D_{im}(T_i, W_i^{-1}(S_i))$.

[1] In practice, when working on image i constructing the model from $N-1$ other images tends to give more generalisable models and lead to faster convergence.

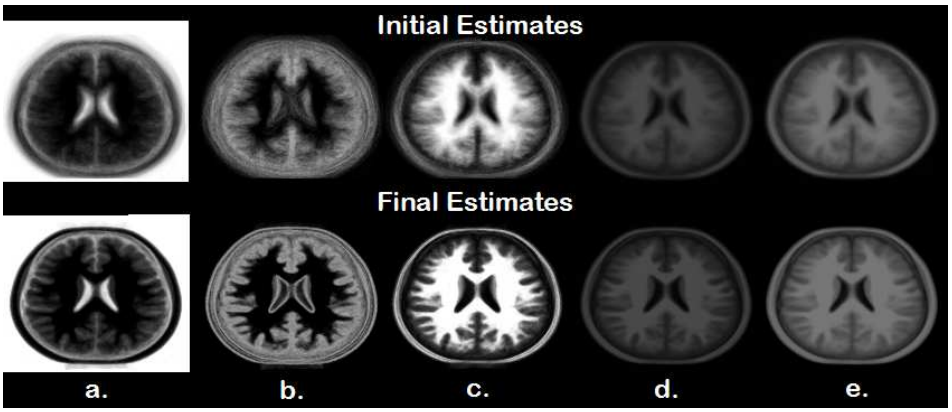

Figure 3: Results of analysis of brain images: a.-c. composition model estimates for the three tissue classes (components), d. and e. reconstructed images of two training set examples (reference frame)

3 Results

We applied the proposed method to two sets of medical images of different anatomies taken with different imaging modalities, a set of 28 X-ray digital radiography (DR) images of the knee joint and of a set of 37 near equivalent 2D slices of magnetic resonance (MR) images of the brain (Figure 3) [2]. For the X-ray images we adopted an absorption IDM which has 2 classes (no radiation and full radiation) at extremes of the intensity range represented as delta Diracs and all intensities in between are considered fractional. In addition we used sum of absolute differences for both the image similarity, $D_{im}()$, and Bhattacharya distance as the divergence between intensity distributions, $D_p()$.

Figure 3 a.-c. shows composition models for the three (tissue) components present in the MR brain images processed by the proposed method. In the final estimates all three classes are crisply delineated and in close agreement with the anatomical distribution of white and gray matter (WM, GM) and CSF in the human brain. Structure reconstruction images S_i corresponding two different training set examples are shown in Figure 3 d. and e. It can be seen that their intensity statistics have been reproduced faithfully by the algorithm. In both cases, the composition model starts from a very fuzzy estimate and becomes more accurate as the alignment across the training set examples is established.

Figure 4 shows the results of analysis on the knee X-rays. This is a difficult data set containing projections of a structure with highly unconstrained pose, scale and image statistics, see 4e. Groupwise intensity registration [5] fails to converge resulting in a mean image 4b, very much like the mean of the non-aligned set 4a. The proposed approach however converges and its mean 4c. clearly shows the main structures. Final absorption (composition) model is shown alongside in 4d. Final deformation fields for three different examples produced by the proposed algorithm are shown in 4e. They demonstrate its ability to deal with large variations in pose and intensities robustly and converge despite the fact that some examples have diverged during affine registration (final example). These failures are caused by the generally sparse structure of these images failing to con-

[2] David Kennedy of the Center for Morphometric Analysis, Boston, provided the MR and Visaris d.o.o. provided the DR imagery

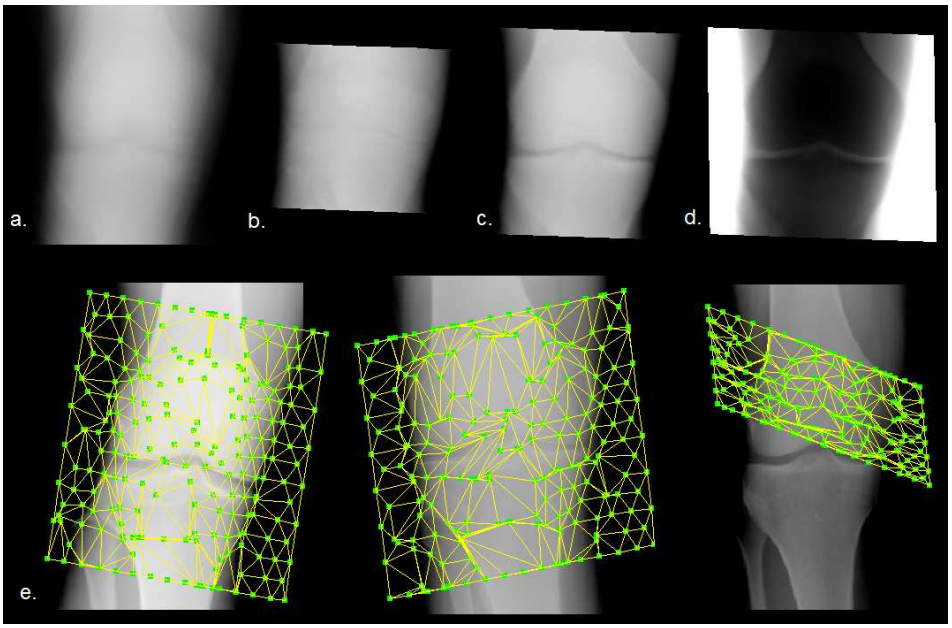

Figure 4: Results of automated analysis of knee images: a. initial mean, b. mean derived using groupwise intensity registration, c. mean derived using the proposed approach and associated composition model for full radiation d., e. final deformation fields for three different images using the proposed approach

strain a powerful global search such as affine registration and could be corrected using relatively straightforward regularisation across the set.

Quantitative evaluation was performed on WM, GM and ventricle (CSF) labels manually defined by experts on the MR brain data using a Tanimoto overlap based metric proposed in [6] (no such ground truth was available for the knee images). The metric measures fuzzy overlap of segmented regions between all pairs of registered images in the set. Results for inverse volume normalised ($TO_{IVol-All}$) [6] and mean of pairwise overlaps for individual as well as all labels (TO_{Label}) are in Table 1. The proposed automated analysis framework (AAF) system was compared to i) pairwise registration where each image in the set is registered to a common reference image selected either randomly PW-random or one closest to the mean of the set PW-opt, ii) groupwise registration where the set is registered to its progressively sharper intensity mean [5] (all using 24x24 point piece-wise affine deformation field and sum of absolute differences objective function) and iii) fluid flow registration (Fluid) [3], using a dense deformation field (defined at each pixel), sum-squared difference objective function, viscosity coefficients $\lambda = 1$ and $\mu = 500$, tolerance for convergence 1e-3, two levels of scale and time step selected by Brent minimization.

Table 1 shows that the proposed algorithm outperforms other systems for all metrics and labels. Figure 5a. shows these results graphically (TO_{all}) including measurement errorbars as well as final intensity means for the PW-opt, GW and the proposed approaches in comparison to the initial mean. Also shown on Figure 5a. as the dashed line is the $TO_{All} = 0.717$ level obtained for groupwise registration of label images, in a way establishing an upper limit on the performance for the chosen registration approach (de-

Metric	PW-rand	PW-opt	GW	Fluid	AAF
$TO_{IVol-All}$	0.591	0.61	0.646	0.651	**0.69**
TO_{All}	0.603	0.616	0.652	0.635	**0.693**
TO_{WM}	0.662	0.664	0.696	0.684	**0.747**
TO_{GM}	0.551	0.537	0.59	0.578	**0.633**
$TO_{Ventricle}$	0.596	0.664	0.669	0.685	**0.69**

Table 1: Quantitative label overlap scores for registration results of various approaches applied to the MR brain data (best score given in bold)

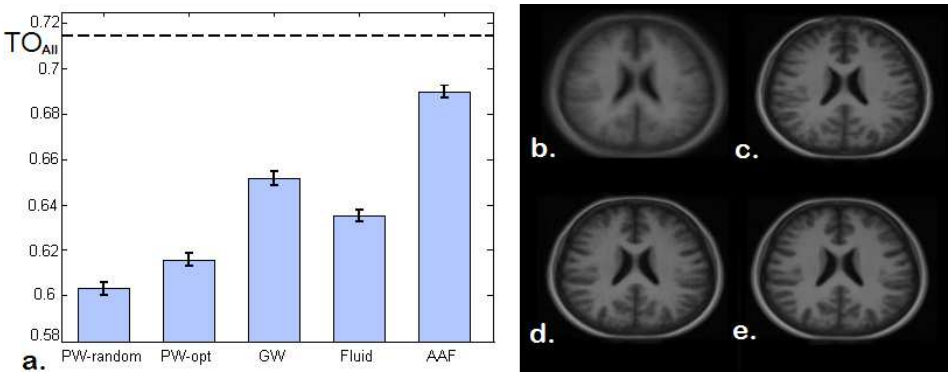

Figure 5: MR brain analysis resuts: a. Label overlap (TO_{All}) results for various approaches, b. initial (non-aligned) mean intensity of the images and c-e. final intensities for the PW-opt, GW and proposed AAF approaches

formation field representation and optimisation scheme). The proposed method achieves overlaps only 2% lower than this limit and much closer than any of the other methods using the same registration approach (in comparison the equivalent reference value for the tested fluid registration approach is $TO_{All} = 0.672$).

4 Discussion

We have demonstrated a powerful algorithm for automated analysis of deformable structure in groups of images. By constructing a model of structure composition, rather than intensities, we decouple the model from details of the imaging process, and concentrate on explicitly learning object structure. The system should be capable of registering images from different modalities. In evaluations on two challenging datasets the proposed framework outperforms other state-of-the-art approaches, despite relying on relatively simple intensity models for segmentation and a relatively coarse deformation field representation.

Future work will include a full implementation to deal with full 3D structures (the extension is natural) and exploring robust segmentation that includes spatial as well as local gradient information. Further consideration will also be given to automating he optimal choice of intensity models for a given dataset, using approaches such as MDL [10, 11] as well as derivation of generic models capable of dealing with various types of objects and image data.

References

[1] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert. Consistent group-wise non-rigid registration for atlas construction. *Proceedings of the IEEE Symposium on Biomedical Imaging (ISBI)*, pages 908–911, 2004.

[2] D. Rueckert C. Xiaohua, M. Brady. Simultaneous segmentation and registration for medical image. In *Proceedings of MICCAI 2004*, number 3216 in Lecture Notes in Computer Science, pages 663 – 670, 2004.

[3] G Christensen, R Rabbitt, and M Miller. Deformable templates using large deformation kinematics. *IEEE Trans. Medical Imaging*, 5:1435 – 1447, 1996.

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models – their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] T.F. Cootes, C.J. Twining, V. Petrovic, R. Schestowitz, and C.J. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In *Proceedings of the* 16^{th} *British Machine Vision Conference (BMVC)*, volume 2, pages 879–888, 2005.

[6] W Crum, O Camara, and D Hill. Generalised overlap measures for evaluation and validation in medical image analysis. *IEEE Trans. Medical Imaging*, 25:1451 – 1461, 2006.

[7] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modelling. *IEEE Transactions on Medical Imaging*, 21(9):1151–1166, 2002.

[8] N. A. Thacker M. Pokric and A. Jackson. The importance of partial voluming in multi-dimensional medical image segmentation. In *Proceedings of Information Processing in Medical Imaging (IPMI)*, volume 2208 of *Lecture Notes in Computer Science*, pages 1293–1294. Springer, 2001.

[9] D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3D statistical deformation models using non-rigid registration. *Lecture Notes in Computer Science*, 2208:77–84, 2001.

[10] C. J. Twining, T. F. Cootes, S. Marsland, V. S. Petrovic, R. S. Schestowitz, and C. Taylor. A unified information-theoretic approach to groupwise non-rigid registration and model building. In *Proceedings of Information Processing in Medical Imaging (IPMI)*, volume 3565 of *Lecture Notes in Computer Science*, pages 1–14. Springer, 2005.

[11] C. Twining V. Petrović, T. Cootes and C. Taylor. Automatic framework for medical image registration, segmentation and modeling. In *Proceedings of Medical Image Understandind and Analysis (MIUA)*, pages 141–145, 2006.

[12] A. Yezzi, L. Zollei, and T. Kapur. A variational framework for inregrating segmentation and registration through active contours. *Medical Image Analysis*, 7:171–185, 2003.

[13] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20:45 – 57, 2001.

[14] Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977 – 1000, 2003.