Groupwise Construction of Appearance Models using Piece-wise Affine Deformations

V. Petrović, T. Cootes, C. Twining, R. Schestowitz and C. Taylor

Imaging Science and Biomedical Engineering, The University of Manchester, M13 9PT

1 Introduction

This work explores a new technique for automatically constructing statistical models of shape and appearance from groups of images of a structure of interest (e.g. MR images of the brain). By computing spatial and structural correspondences across the images in the group, essentially registering them, we can easily build such a model. A widely used approach to register the images is to select a reference example and find correspondences between this and all other images in turn. An alternative approach adopted here is to optimise a dense correspondence field accross all examples in a groupwise manner, using minimum description length, of the data encoding by a model constructed from the current correspondence, as an objective function. The correspondence is defined using a piece-wise affine interpolation between a set of control points placed on each image. We show that the proposed approach is valid for both two and three dimensional modelling and briefly consider the problem of the optimal choice of control (landmark) points for model construction and its ability to provide a broader understanding of the imaged structures.

2 Method

We model the appearance of a set of images by deriving linear models of shape and texture variation accross the set. We will assume that shape of each example in the group is uniquely defined by the position of a set of control points for which we model as $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb} + \delta\mathbf{x}$ where $\bar{\mathbf{x}}$ is their mean position accross the set, $\mathbf{P}$ defines a set of modes with shape parameters $\mathbf{b}$, and $\delta\mathbf{x}$ is a set of residual displacements (to the true locations). Texture information, sampled by warping each of the examples into the reference frame using the deformation field defined by the control points is also modelled with a linear model.

Any training set image could be encoded and reconstructed by the model given the (shape and texture) model parameters and residuals. The basic idea is that better models will produce a more compact encoding and we treat model construction as an optimisation problem in which the objective function is the description length of the model encoding of the training set. To derive the description length we must consider the cost of all the data required to reconstruct the training set: texture and shape parameters, residual displacements $\delta\mathbf{x}$ and the set of residual image values (one per pixel). In an optimal coding scheme which is assumed here the cost is related to the log probability of each term. The total cost of the training set also includes the cost of transmitting the model: shape and texture mean vectors and shape and texture mode matrices.

The first step in model construction is the groupwise registration that derives the dense correspondence between training set images and proceeds as follows:

1. Select one image to be used as an initial reference and select control points on this reference image
2. For each image in turn estimate movement of control points to optimise match to the reference image
3. Groupwise registration:
   (a) For each image in turn, compute shape and texture models from point positions and intensities of all other images, compute the best-fit of the model to the current target image and modify its control point positions to minimise the cost of encoding it using the model.
   (b) Repeat until the process converges or an objective aim is reached.

During registration, the objective function depends only on the location of the control points $\mathbf{x}$ which are optimised in a coarse to fine regime. This involves manipulating coarser grids with fewer parameters which move sets of control points $\mathbf{x}$ smoothly. The finest stage of the search optimises the values of $\mathbf{x}$ directly. At each stage a simple downhill gradient search of the objective function is used as the optimisation method. At each groupwise stage the control points are essentially aligned to the leave one out mean of all the others (model mean) which gets progressively crisper and more accurate as the registration progresses.

An important issue for registration is the choice of a representation for the deformation field that essentially defines the dense correspondence between the images. The algorithm above requires us to be able to invert the deformation efficiently and we use a piece-wise linear representation, in which the region of interest is mosaiced by a set of triangles (in 2D) or quadrahedra (in 3D) with vertices being the control points. Within each region we use affine approximations of the deformation field which are easily inverted and although the resulting representation is not smooth (derivatives not defined at the boundaries) we have found it to produce good results and also simple to add constraints to prevent non-invertable mappings. Initial tesselation of the region of interest can be obtained using the Delaunay algorithm. In 3D however care must be taken as a unit cube tesselation has a number of possible outcomes and can cause self intersecting meshes.

3 Results

The described method was tested in two and three dimensions on 104 transaxial 2D MR slices and a test sample of 10 3D MR brains. In both cases registration using an arbitrary choice for the initial reference was performed and appearance models constructed according to the described algorithm. 2D models were evaluated using the concept of model *specificity*, recently shown to be a good surrogate for evaluating correspondence errors, computed as the mean minimum shuffle distance between a set of e.g. $N = 1000$ synthetic images drawn from appearance model and the training set examples (using known affine transformation between the images). The results of 2D evaluation for three different registration approaches within the proposed model construction framework are shown on Figure 1. For the full range of model compactness groupwise approaches outperform pairwise registration of each training image onto a single chosen reference. The results also show a clear advantage of measuring the texture residuals in the reference frame rather than projecting the reference image into each target image frame.

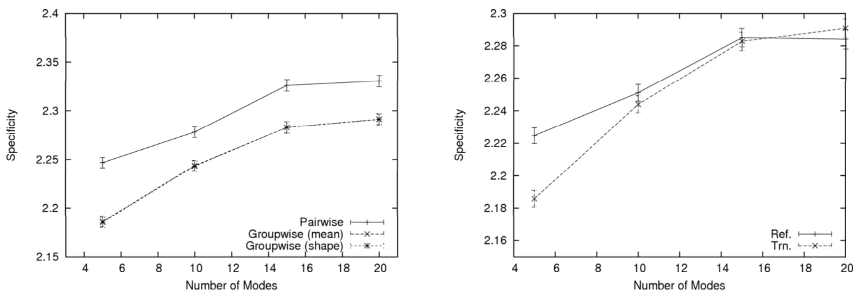

Figure 1. Model specificity vs. registration approach (left) and vs. texture evaluation frame

Model specificity is less practical for the evaluation of 3D models as 3D shuffle distance computation becomes extremely complex. 3D registration is demonstrated in Figure 2 on a pair of corresponding slices of two 3D brain images being registered. The capability of 3D registration is clearly shown as structures from other slices of the target image slide into alignement with the reference slice.

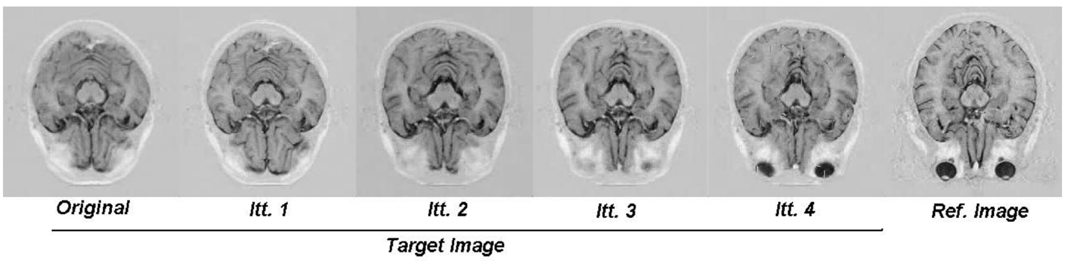

Figure 2. 3D registration

4 Further Work

An important effect on the resulting model quality is the choice of control points. Historically, control points have been chosen *a-priori* so as to facilitate efficient registration. However an *ad hoc* spatial distribution might not be an optimal description of the underlying image structure. By applying criteria analogous to those used in registration (e.g. MDL) and using the obtained dense correspondence between the images it should be possible to optimise the location and tesselation of the control points consistently accross the training set examples and obtain an improved model of the data. Preliminary results show an encouraging improvement in model quality and understanding of the data within the training set and will be explored further.