LITERATURE REPORT

R. S. SCHESTOWITZ
IMAGING SCIENCE AND BIOMEDICAL ENGINEERING
STOPFORD BUILDING
UNIVERSITY OF MANCHESTER
UNITED KINGDOM

ABSTRACT. Construction of robust and concise statistical models is a goal which is difficult to achieve due to the high computational requirements it typically imposes. Numerous attempts have been made to automate the construction of good active appearance models, but none has yet been successful. Potential is born, however, in the unification of model construction and image registration. The evaluation of non-rigid registration, which is based on non-linear warps, has been another issue of great difficulties and automatic selection of good warps is far from trivial.

In active appearance models the main problem is the inability to select good landmarks without human judgement, as well as the difficulty in location and annotation of these landmarks using brute-force only. Non-rigid registration is a quickly emerging technique that can be used to warp multiple images and give some group-wise optimal model, as opposed to a model derived from pair-wise registration that depends on an arbitrary choice of a reference image.

These arguments highlight the benefits summoned by the combination of these two techniques – active appearance model can aid the selection of good warps in non-rigid registration and the functionality of non-rigid registration can help obtain more compact and robust models of appearance. This report outlines the current work in the field as well as some concepts that bear potential or whose realisation can contribute to future endeavours. It critically surveys some existing work and techniques along with the results they have produced. Lastly, this report attempts to identify some existing gaps where substantial improvements can be still made and describe the research planned with its aims and intermediate milestones.

Date: December 16th, 2003.

CONTENTS

1. INTRODUCTION

The work invested in two separate yet related fields calls for a strategic merger that takes advantage of the best of both. These fields are statistical models of appearance and non-rigid registration whose wide-spread use consequently made them independently usable and powerful. As they deal with problems that have a great deal in common, attempts have been made and still are being made to discover how one field can benefit the other and to what extent.

The broad field of image registration encircles some important techniques that academical and clinical research groups have an interest in [1, 2, 3, 4, 5]. An evident rise can be shown in the number of papers published in the field, medical context being a noticeable focus. I turns out that registration is in many respects valid to biomedical data as later discussions stress.

Registration is concerned with the assembly of data which is taken either at different points in time or at some arbitrary time instances where changes due to the passage of time can be ignored. Registration sees the most use in scenarios where multiple *different* objects or subjects are being scanned or where the acquisition method varies. In the case of medical imaging, registration is commonly mentioned in one of three distinct circumstances: intra-subject registration, inter-subject registration and multi-modality imaging. This corresponds to the investigation of changes in one specific subject over time, the investigation and comparison between more than one subject and the fusion of data acquired from different modalities (e.g. CT, PET and MR) respectively.

The main problem that registration is set to overcome is the *alignment* of several images to with the aim of achieving better *correspondence* across the entire set of images to be dealt with. This correspondence can be evaluated by similarity measures, examples of which are given later. With suitable overlap of some given object[1] within a group of images, segmentation, analysis and comparison are significantly more constructive; these are usually impossible guarantee in the absence of that overlap. Correspondence is not always simple to achieve algebraically since the object inspected or the aperture[2] may change position and angle over time or acquisition venue.

[1]The word "object" will from here onwards refer to a structure of interest in n-dimensional space.

[2]In the case of medical imaging, there are even more factors to be considered, as opposed to a camera's aperture.

In reality, additional unwanted effects such as noise, distortion and change in form must be carefully accounted for. In some real-world applications, biological being an ideal exemplar, variability must be handled sensibly to understand the changing structures that are present in an image. Therefore, the correspondence expectations, as well as the permissible degree of freedom, must not be excessively rigid[3]. It is important to ensure that the chosen analysis mechanism caters for some level of flexibility to enable a robust registration process that is immune to high levels of misalignment.

The problem of registration would have been rather simple if it were not for the innate changes that are an integral part of any biological element. Simple alignment is therefore not necessarily sufficient to give good a solution – that is – plausible correspondence. As explained in Section 3 of this report, registration methods cab be further broken down into different types, but their aims remain the same in essence. The methods aspire to find some correlation between two or more images, in which case a new entity is obtained that expresses some informative relations between the distinct images.

Image registration is believed to be capable of positively affecting the performance of statistical models and vice versa. More compact models of variability can be constructed if registration procedures are applied to its training data (confer [6] for more details on learning and training). This is obvious because registration clearly minimises the variability seen. The later sections explain in greater depth how the two techniques come together and how one can be incorporated in the other, whereas the earlier sections attempt to explain and show the commonality between the two.

In some past work, the formation of appearance models, based on registered images provided some fair indication of how desirable that prior registration was. However, the process was slow and there is a need to find better ways of using the two techniques in a cunning and hence more efficient manner.

Quite broadly and even wishfully, upcoming research intends to bring together all phases of the handling of an image, from the moment when images are registered to the point where these are coupled with an appropriate statistical model (and even get segmented and measured). Arguably , it would not be too optimistic to state that model fitting, shape analysis, non-rigid registration, feature detection and segmentation can and should

[3]"Rigid" refers to constrained variability and low model generalisability as explained later. It is significantly different from the term "rigid" in the context of registration.

be put under one *single framework*. In this way, more compact and powerful representation of images can be used – images can be described by the parameters of the non-rigid transforms that ought to generate them. This is in fact what makes this unification of several methods quite appealing when compared with stand-alone active appearance models.

2. ACTIVE APPEARANCE MODELS

2.1. **The Problem.** Image analysis is a general problem that can be tackled in various ways. This analysis is fundamental and essential to many processes such as industrial inspection, motion analysis, face recognition and medical image understanding. What makes this problem intrinsically hard is the inability to take into account single pixels independently to infer the structure they form together. The goal of such analysis is not only to solve the problem correctly, but also to do so efficiently, in a way that is not overly affected by the size of the image, i.e. the scale of the problem.

Analysis often involves *measurement* of meaningful structures in an image and possibly some explanation regarding the *form* of these structures. In order to derive any useful information about a particular meaningful structure , image *segmentation* must first take place. Segmentation is concerned with the identification of certain regions of interest which may be characterised as belonging to the same object. By deriving to image into such regions, understanding of the nature of its constituent components can be gained.

This report concentrates on a top-down approach to analysis. This approach relies on a high-level abstraction of the visual attributes of one structure. Alternatively, and often more usefully, this abstraction can represent a *collection* of structures that together form another structure. The reason why such an approach is referred to as a top-down approach is that it bears some existing information that it attempts to *fit* to the problem posed. It makes assumptions about the problem and is in some sense taking a preliminary overview on the structures in an image.

The rest of this section will describe popular methods of top-down image analysis, but will focus on active appearance models on the expense of other, less relevant methods.

2.2. **Statistical Models.** Active appearance models are somewhat of an extension to active shape models and a brief introduction to shape models may be worthwhile. Given a collection of images that depict an object which possesses some innate properties, it is then possible to express the visual appearance or shape of that model in a way that discards subtle changes in view-point , object position, object size and is robust to some level of object deformation. That object which appears in the group of images need not even be the exact same one; it can be an object belonging to one common *class*. Some variation can be handled reliably by simple transformations, but their functionality is inevitably very limited. There are statistical means which allow the encoding of the variability as was *learned* throughout a so-called training process. That training process does not require far more than an exhaustive inspection of the set of images. However, in order to interpret several image, some simplification steps are required as the images are expected to be relatively large in practice – certainly large enough to result in an exponential blow-up[4]. A method is sought which reduces the amount of information that is required to describe the object of interest and the different forms it can take. In most cases, edge detection is sufficient to capture regions or points of greater significance in the image[5]. Such points are often chosen to become what is entitled *landmarks*. Landmarks are positions in the image which effectively distinguish one object from another in the set of images. They also have a some interesting spatial traits and their low proximity can form near-optimal curves (or contours) which together make up *shapes*. The concatenation of the coordinates of these landmarks can then describe an image (or rather the object focused on) in a by concise and useful representation. In 2-D, for n landmarks, a vector of size $2n$ can roughly infer the shape of the object present in an image. For the naíve pixel-wise representation, not only will a space of $width$x $height$ need to be allocated, but also the manipulation process on this data will slow down considerably.

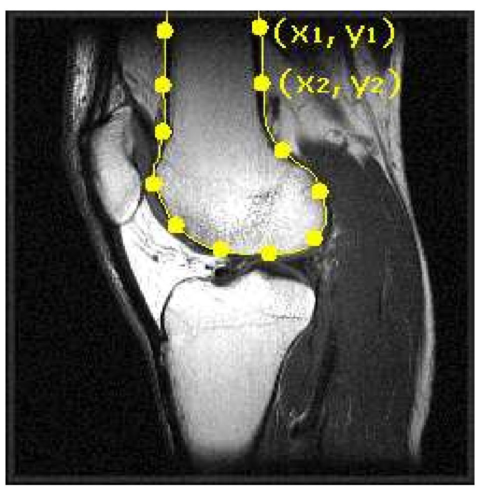


Figure 1: Landmark identification

[4]Current model-based methods typically deal with only the order of tens of thousands of pixels. High-resolution medical images can contain millions of pixels.

[5]Edges and corners usually hold more information of use for subsequent analysis and aid segmentation. They lead to better identification of the different objects residing in the image.

With the concise landmark-based representation as above set to be the convention and a collection of fair-sized vectors rather than a collection of images, it should be possible to express (in a feasible way) the legal range of each one of the vector components. This in essence establishes the *model*. It is an entity that can be manipulated to reconstruct all the images (or objects) it originated from and far beyond that. This model encapsulates the variation which was learned from the data and it usually improves its performance as more legal examples are viewed and "fed" to support some further training. Alternation of the parameters of the model can generate new (unseen) examples as long as that value alteration remains loyal to the legal range, as learned from the training examples. The vector representation mentioned beforehand can be also looked at as a description of some fixed location in space that comprises $2d$ dimensions (see illustrative scatter below). This turns out to be a useful analogy as will be seen later when dimensionality reduction is applied.

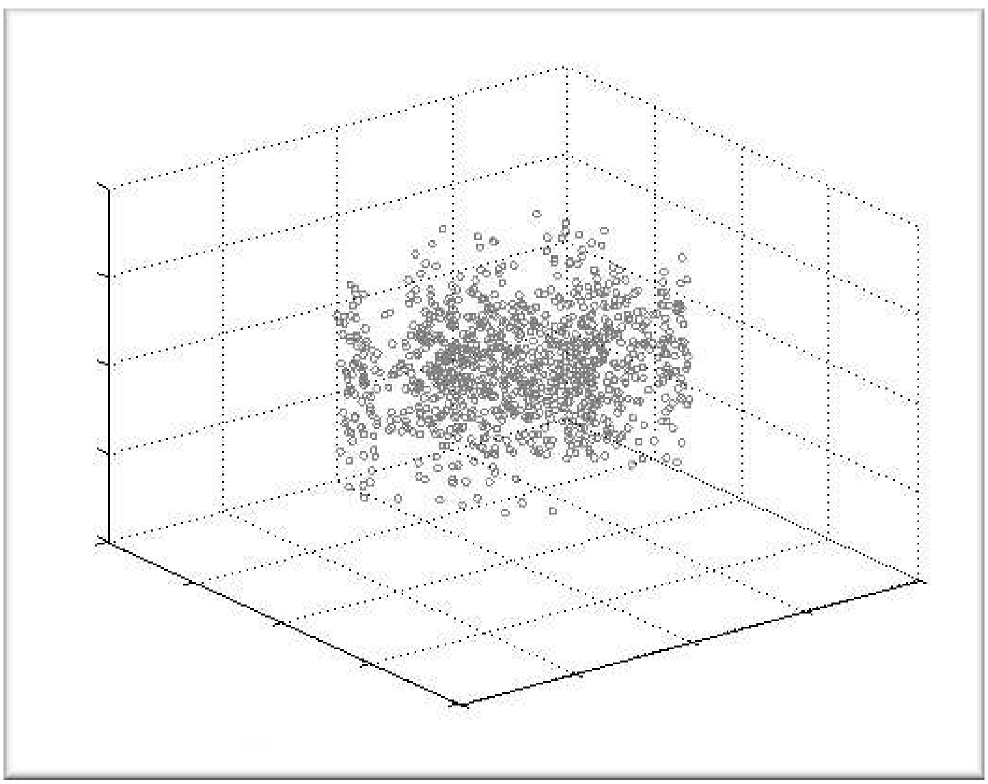


Figure 2: 3-D Scatter

Shape models are “statistical information containers” which can be built from the images with overlaid landmark points identified and recorded. A common grid whose purpose is to ease collective alignment is a crucial way of achieving consistency amongst the coordinates of all landmarks. Some more issues that are concerned with normalisation steps are described slightly more accurately later in this document. A human expert usually performs annotation or landmarking of the images with the aid of some computer-assistant tools. In recent years, alternatives which are automatic showed great promise [7] and these extend to 3-D too [8].

Active appearance models were later developed by Edwards *et al.* [9, 10] and the great advantage of these was that they were able to sample grey-level data (Stegmann *et al.* [11] incorporate full colour by now) from images rather than just points. Therefore, they held information about what an image *looks* like rather than just its *form* as visualised by contours (or surfaces in 3-D). Just as points in the image were earlier chosen, grey-level values (also referred to as *intensity* or *texture*) could be systematically extracted from a normalised image and put in an intensity vector. This normalisation process and the representation of this intensity vector is described later in this section.

What enables appearance models to exhibit quite an astonishing graphical resemblance to reality is that at the later stages a *combined* vector is made available. It incorporates *both* shape and intensity and is aware of how a change in one affects the other. Hence it has a notion of the *correlation* between the two – a notion that is dependent on the training data and Principal Component Analysis. Although appearance models are not as quick and accurate as shape models, they contain all information that is held in the shape models and in that sense are a superset of shape models. Also, some techniques have been developed and employed to speed up active appearance models. Tasks such as the matching of an appearance model to some target image are described later in this section and illustrated in [12].

2.3. **Appearance Model Construction.** The first step is concerned with the establishment of a model that not only describes a *mean* form of some object in an image, but also the legal variation that can be applied to the mean in order to create new legal object instances. A model formulates the form that vectors can take and that vector can easily be translated to a visual description. More desirable models will not be excessively lenient and should allow recognition and acceptance of only reasonable variations of the object under investigation. There is a convenient mathematical way of expressing this variation and that is to assign some parameters to each mode of variation[6]. When change in these parameters occurs and the mean shape is deformed accordingly, there will be a direct effect on the appearance of the result, where that result is really just a simple vector. Rather usefully, each legal instance can always be uniquely and fully described by the parameters which were used to generate it from the model. The shape and vector representations are equivalent and interchange. Visualising results is often convenient visually while logical operations are better thought of in terms of vectors.

To begin encoding the form of an object, landmarks need to be identified and statistical analysis applied to express these spatial shape properties, namely the landmark coordinates. From this analysis, a mean shape is obtained and it can be denoted by $\mathbf{x}_{mean}$or $\overline{\mathbf{x}}$. To obtain this mean, the procedure that is commonly used is Procrustes analysis. The generalized Procrustes procedure (or GPA for Generalised Procrustes Analysis) was developed by Gower in 1975 and has been adapted for shape analysis by Goodall in 1991. It processes each component of the vectors derived from the images and returns for each component a value that is said to be the mean. From here onwards, this vector which represents the mean of the data will be referred to as $\overline{\mathbf{x}}$. Each shape $\mathbf{x}$ is then well-formulated by the following:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{P}_s \mathbf{b}_s. \tag{2.1}$$

The matrix $\mathbf{P}$ represents the eigenvectors of the covariance matrix (set of orthogonal modes of variation) and the parameters $\mathbf{b}_s$ control the variation of the shape by altering these modes of variation[7]. Eigen-analysis is used quite extensively in the derivation of the expression above, but it will not be discussed in detail in the remainder of this report. Instead, a short explanation will be given on Principal Component Analysis [13, 14] which from here onwards be referred to as PCA. What is worth emphasising is that the only variant in the model described above is $\mathbf{b}_s$and as the values of $\mathbf{b}_{1<i<s}$ are infinite ($\mathbf{b}_{1<i<s} \in \mathbb{N}$), the same must hold for $\mathbf{x}$. There is an infinite number of shapes, each of which can be generated from one choice of value for each model parameter.

[6]Offsets of standard deviation units from the mean of each mode then illustrate the effect each mode has.

[7]Because there is a total of n modes of variation, $1 < s < n$, i.e. only n parameters exist.

At this stage, each of the images should be aligned to fit a common space. In practice, that space is implicitly defined by the mean shape. Rigid (or Euclidean similarity) transformations, namely translation, scale and rotation, are not always sufficient to warp all images into that common space. Nonetheless, it is crucial that good fitting is obtained before the sampling of grey-level begins. Following these basic transformation which align all images, the displaced control points of each image contain shape-normalised patches. These patches are available for construction of texture vectors and Barycentric arithmetics, which often gets applied to computer graphics and stereo vision, is used to describe the location of all corresponding points. This location is affected by the warps applied to shift a given shape into the space of the mean shape. Triangle meshes are created by stretching lines between neighbouring control points and intensity values are captured one by one (along a chosen grid of points to be samples) and stored in a vector of texture. Each component in such a vector captures the intensity (or colour) of one single pixel as was learned from the examples. Statistical analysis as the one above results in the following formulation for texture:

$$\mathbf{g} = \overline{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g. \tag{2.2}$$

The use of the algorithm above implies that for short vectors and a low number of pixels sampled, noncontinuous appearances will be too easy to spot. In fact, objects will often appear to be nothing more than a collection of polygons that do not quite resemble realistic appearances[8]. To compensate for this, algorithms from the related field of computer vision can be used, e.g. Phong and Gouraud shading. In practical use, geodesic interpolation is used and the results can be quite astounding considering the low dimensionality of the data.

The models above (2.1, 2.2) are expressed both linearly and compactly – a highly desirable and manageable form. This is due to PCA which reduces the length of the vectors describing shape and texture. Although Eigen-analysis is involved in the process, it is not essential for the understanding of PCA which works as follows.

It is possible to visualise the data as points in a high dimensional space as was earlier argued. By placing all images in that space, it is expected that some cloud of points will be present at a specific somewhat confined region.

[8]One of the main aims and great power of appearance models is full synthesis so photo-realism is at a premium.

The breadth of this region or the size of that cloud will depend on the variation amongst the images (or more generally data) that is being visualised. PCA obtains the eigenvectors and eigenvalues of that cloud of points. The highest eigenvalue will correspond to the most significant eigenvector (see the single-headed arrow in Figure 3). It indicates the direction that best distinguishes the image data and is expected to be the longest one too – that is – the one whose magnitude is the greatest. This is in fact what is considered to be the principal component which describes that data. In a recursive manner, at each stage in the process, the principal component is saved and put aside until only negligible components are present. The recursion will therefore deal with simpler, more uniform data as more and more principal components are set aside and leave a data of lower dimensionality. A smaller number of components can then be used to express the variation up to a relatively high level of fidelity. The process is lossy, but so are some other stages in model construction including the choice of a finite number of landmarks. That loss is controlled in the sense that one can choose the minimal amount of variation that must be accounted for[9]. PCA is used to gain speed while retaining the best descriptors of variation and difference in shape and intensity. What this all comes down to is the acquisition of a model that is smaller in size and is easier to deal with.

[9]A common choice is 98% of the observed variation which means 2% of the variation is not accounted for. This 2% of variation is usually the least informative though which is the what PCA is intended to accomplish.

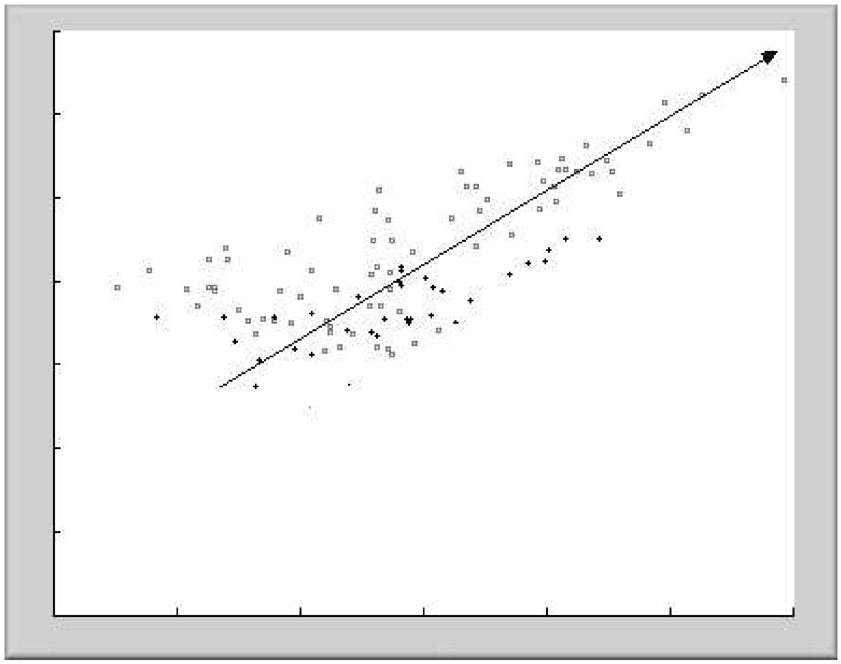


Figure 3: Principal component in 2-D

The two components $\mathbf{x}$ and $\mathbf{g}$ above need to be merged to establish a model that accounts for both types of variability (shape and intensity) and holds within it the correlation between the two.

The parameters $\mathbf{b}_s$and $\mathbf{b}_g$are aggregated to form a single column vector

$$\left\{ \begin{array}{c} \mathbf{b}_s \\ \mathbf{b}_g \end{array} \right\}. \tag{2.3}$$

It is in some sense a simple concatenation of the two. However, since the values of intensity and shape can be quite different in nature and granularity, some weighing is strongly recommended. If no weighing is applied, then the spread of the points in space is quite undesirable and the components to be identified by PCA are not as beneficial as they would be otherwise. If some values are far greater than others, vicinity takes a turn for the worse and the cloud might be elongated instead of nearly spherical (a 3-D analogy). For rather spherical spreads (or those of almost homogeneous variation), a greater number of large components will be available for selection. Consequently, the variation expressed by a fixed and constant number of principal components will be higher.

A weighing matrix that resolves the problem introduced above is by convention named $\mathbf{W}_s$ (s corresponds to *shape* as by default this matrix scales the shape parameters only). The form in which coordinates are stored in $\mathbf{x}$ depends on the accuracy required (e.g. integer and floats), the image size and the number of dimensions whereas for grey values this form is dependent on the number of bits used per pixel[10]. With weighing in place, the aggregation would take a form such as

$$\left\{ \begin{array}{c} \mathbf{W_s b}_s \\ \mathbf{b}_g \end{array} \right\} \tag{2.4}$$

where $\mathbf{W}_s$ is chosen to minimise inconsistencies due to scale. Lastly, by applying a further PCA to the aggregated data, the following combined model is obtained:

$$\begin{array}{c} \mathbf{x}_i = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c}_i \\ \mathbf{g}_i = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c}_i \end{array} . \tag{2.5}$$

The appearance is now purely controlled by the parameters $\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n$ and there is no need to choose values for two families of distinct parameters as before. This combined model has the benefits of the dimensionality reduction performed which is based on shape as well appearance. This means that it now encompasses all the variation learned and the correlation between these two distinct components. Since PCA was applied, the number n of parameters $\mathbf{c}_i$ is expected to be smaller than the number of $\mathbf{b}_s$ and $\mathbf{b}_g$ put together.

2.4. **Model Training.** A descriptive statistical model is now available for utility and various analyses. That model is a type of flexible deformable entity that can describe any instance of object or image[11] in the range of

[10]For colour it is common to use 24 bits and for grey-level just 8 bits. For more compact statistical appearance models, less than 8 bits (256 shades of grey) might suffice to achieve good results and in medical imaging 12 bits are nearly a standard in acquisition.

[11]The distinction here is hard because the model can describe more than one valid independent object and usually represents only a partial section of the entire image. In a medical context, the term *atlas* fits somewhat more nicely and it usually describes a single organ or anatomical structure.

the training set[12]. Assuming that the training set was infinite in size or comprised all possible instances that the model might be presented with, it should then be considered a powerful, fully compatible, flawless model.

It is still not trivial in any case how one should deform the model to achieve an appearance instance that is valid. It is now a completely opposite problem that a user of this model is faced with: how can one model generate new instances after existing instances generated that one model? In some sense, an inverted operation is needed so that the model can be used in the opposite way to the means in which it was created. Things are not very simple in reality and the alteration of model values needs to be guided by some minimisation (the next section elaborates on this) that obtains the matching which is required. Unfortunately, in an expectedly high dimensional space as above, the process is almost endless unless extra knowledge about this minimisation problem is provided in advance.

The way in which this problem can be circumvented quickly involves learning how the parameters $\mathbf{c}_i$ affect the model[13]with respect to a typical target. Each parameter in $\mathbf{c}_i$has an unequalled effect on different regions in the model, e.g. its size, intensities and so on. By changing the value of each such parameter and recording the change that is perceived in an image (using pixel-based comparison of some kind), a type of deformation index can be maintained. This index indicates which parameters should be changed and if so in what way at to what degree in order to approach good overlap between a model and some target image.

More formally, the procedure works as follows:

For the model parameters $\mathbf{c}_i$ where $1 < i < n$, a parameter change $\delta\mathbf{c}$ (where one parameter value or more can be readjusted) is applied to generate some new shape and texture. $\delta\mathbf{c}$ expresses in a vector-based representation the offsets that each of the original parameters $\mathbf{c}_i$ is subjected to. The exhaustive pixel-wise difference in intensity[14]is calculated in accordance with:

[12]There is a subtlety which makes this phrasing a bit deceiving and inaccurate. The word "range" is a gross terminologically equivalent for the area that stretches in between the space of training set instances. It can be conceived as the space defined by a Gaussian distribution cloud that is deduced from the training set.

[13]There are some more complex considerations as the model needs to be aligned properly as well as change form.

[14]A simple raster scan that account for all pixels should clearly be fast under most contemporary computer architectures. This is indeed the case if simple operations like subtractions are pipelined on the ALU (arithmetic and logic unit).

$$\delta\mathbf{I} = \mathbf{I}_{model} - \mathbf{I}_{image} \tag{2.6}$$

to produce a new vector of intensities (the differences). This vector can also be visualised to display this difference to a human eye. A simple measure of difference is used although this need not necessarily be the case. Sum-of-squares of the pixel differences is then used because larger quadratic differences will have a greater effect on the final measure and summation then only consists of positive values. For example:

$$\delta\mathbf{I} = sumofsquares(\{-1, 3, 5, 2, 6, -10, -1\}) \tag{2.7}$$

then becomes

$$\delta\mathbf{I} = sum(\{1, 9, 25, 4, 36, 100, 1\}) = 176 \tag{2.8}$$

as opposed to

$$\delta\mathbf{I} = sum(\{-1, 3, 5, 2, 6, -10, -1\}) = 4. \tag{2.9}$$

With this measure of intensity difference recorded, a correlation can be expressed between the parameter change and this difference as it appears in image space where a model is superimposed on some target. A target image which is the model in its mean form is needed here to be used for basic comparison. This quantitative measure of difference obtained will however indicate solely the "goodness" of the parameter change and not the overall effect that it has on the image. This means that it will not necessarily be obvious what parts in the two entities (model and target) remained similar and which ones did not[15]. A type of a sequential data such as a vector is hence more useful as it retains the location of each computed difference value. Unsurprisingly, this consumes far more space. In either case, under the premise that space is more expendable than time complexity a vector of difference is calculated and the correlation can be formulated as follows:

[15]The vector's distribution of values, i.e. positions with high absolute values, can answer this question quite coarsely.

$$\mathbf{c}_i \rightarrow \mathbf{c}_i + \delta\mathbf{c} \rightarrow \delta\mathbf{I} \tag{2.10}$$

This type of offset $\delta\mathbf{c}$ that was applied to the collection of parameters $\mathbf{c}_i$is accompanied by a global change in intensity values across the image frame. This correlation can now be stored aside and become accessible from an index as its size is proportional to the image size. The storage is dictated by the following (somewhat artificial) relation:

$$\delta\mathbf{c} = \mathbf{A}\delta\mathbf{I} \tag{2.11}$$

where $\mathbf{A}$ is a matrix recording the change in intensities due to the parameter change $\delta\mathbf{c}$. This is a type of matrix which is analogous to an *n*-dimensional vector that expresses the change which was discovered off-line. It linearly defines (in a possibly high dimensional space) the linear relation between change to the parameters and change to the intensities, or more precisely the different image. It can be looked up directly later on when performing a search and thereby avoid re-computation in a virtually recurring and almost identical problem.

The most fundamental and perhaps even compact procedure will carry out the steps above for each of the modes of variation, as well as the basic geometrical linear transformations. This can be a very laborious and cumbersome process although it depends on the robustness prescribed. As the next stage illustrates, models that are not rich enough will fail to converge in difficult scenarios, a classic example of which is inappropriate initialisation.

The matrix A holds real valued numbers (preferably of limited accuracy to decrease space requirements and access speed). The values in this matrix form a beneficent map that guides exploration for good parameter changes; this will be of great use when fitting the model to a target. In practice, such matrices are visualised by showing negative values as dark colour and positive one as increasingly brighter values.

2.5. **Searching and Fitting.** The final stage which is arguably the most fascinating one involves the use of the model above, as well as the correlations learned and recorded for that model. It is possible to carry out a search which is driven by the difference calculated between the model and a given target image. In practical terms this means that fitting of the existing model will slowly be improved until the model approximately covers the target[16]. The model state then holds (in the form of parameter values) some information about the target image and this information can be further analysed.

The search is reliant on error (or conversely similarity) measures which are repeatedly calculated after each attempted parameterisation of the model. Each such change in parameter values is primarily guided by the matrices that express the correlation between variation modes (the Similarity transformations as well as modes of appearance change) and the intensity values.

The model, as shown in Figure 4, is initially placed somewhere inside the image frame, with reasonable proximity to its target. If the model is placed too far from its target, there is a danger that it will not be able to converge to the target correctly. It will most likely get stuck in a local minimum and the outcome can be severe in a more crucial application such as medical imaging (or perhaps more drastically, computer-guided or -aided surgery). The reason why good initialisation is essential is that significantly large displacements are rarely learned off-line and the difference between the target and the model is quite meaningless unless there is at least some partial overlap or commonality.

The algorithm which is used to perform the search sensibly has a general form that resembles the following:

- Place the appearance model somewhere in the image, preferably at the centre where the target of interest is likely to lie[17].
- For the appearance model in its current state and the static target do:
  - Calculate the difference between the model and the target.
  - Using the correlations learned off-line[18], set new values for the parameters $\mathbf{c}_i$.
  - Compute the new difference measure between the model and the target.

---

[16] This process of fitting strives to converge to the global minimum (of difference measure). Realistically speaking , the model and the target never reach complete equivalence, namely the difference value of 0. Even if the target was used to train the model, PCA would corrupt the connection between the two.

[17] Advanced knowledge about the problem is highly helpful at this stage, otherwise some bottom-up image analysis is a must.

[18] If these are not available, some guessing would be an alternative. It is important, however, to learn from the experience gained during this independent run of the program or else the optimisation would behave senselessly and lead to improvements very slowly. General optimisers ought to make a good judgement as such.

    * Save the new state of the appearance model if the difference has been lowered, i.e. similarity is being approached.
    * If not, try re-adjusting the parameter change, potentially with inclusion of a scaling coefficient $k = 1.5, 0.5, 0.25$ and so on. This often achieves good results, although it is a heuristics-driven technique.
- Iterate while no convergence has been reached and improvements are still observed at times.

More advanced methodologies and algorithms are used at present, but better clarity is achieved by adhering to simplicity.

Figure 4: Model and target fitting

The technique of matching an appearance model to a target image is well-depicted by a staged simulation, a video clip or a large sequence of images as in Figure 4 above. Surprisingly, only a few dozens of iterations are required in order to get good results. This of course depends on the algorithm, the magnitude of the problem and its internal intricacies, e.g. fitting a perfectly round ball versus a human hand.

2.6. **Real-time Active Appearance Models.** Some machines are able to deal with small-sized fitting in real-time [15, 16]. It is possible to track faces in a video (frame rate should then typically be 24 frames/second and 15 at the minimum), but the resolution catered for is often relatively low (less than 100x100 pixel). Applications that respond so quickly were made far more practical owing to a multi-resolution (multi-scale) approach. In order to decrease the total run-time, varying increasing image resolutions become available for selection at each search iteration. Finer resolution images are usually used at the later stages of the search, whereas low-resolution (coarse) ones at the very start. Since the similarity between the model and the target is poor at the start, the resolution (and hence the scale of the objects) will have little effect on the fitting. Some visual examples of AAM search are shown in [12].

2.7. **Other Applications.** A very common use of AAM's is for medical image analysis and face recognition. Active appearance models possess traits that make them robust and effective in the biological domain, whereas industrial inspection, for example, presents some inherently different problems. These problems are often solved more quickly by other approaches that are based on lower-level knowledge about the image contents. Since a broad range of tasks are performed in industrial inspection, however, it is also valid to assume that the suitability of top-down approach is irrespective of the problem.

In order to visualise biological shapes and full appearances, a model which handles anatomical variability and change needs to be used. It must account for natural or pathological changes such as the change in form of organs. Greater variability can be encountered when aligned images are obtained from different subjects in a population (inter-subject), the same subject at different time instances (or different sites) or when having to account for movement such as the that which occurs due to respiration, the cardiac cycle, etc. A separate case to consider is multi-modal imaging which will not be explained in any detail although it is a fast-developing area.

For a good overview on many of the different image analysis techniques, Sonka [**?**] is a valuable source. For a good review of model-based image analysis, papers from Cootes *et al.* are an even better source although the ideas are partially based on concepts such as snakes, bending-energy and active contour models that led to the development of active shape models. Sonka discusses all these in depth.

2.8. **Existing Extensions.** Wavelet compression techniques are used to reduce the troublesome space requirements (especially in 3-D) and make active appearance models far more compact. There are also some application-specific extensions such as the implementation of view-based models [18] and coupled-view models [19] for faces. The principal idea is that 5 different models can express full appearance irrespective of the wide range of viewing angles around the head. Due to the symmetry of a human head, only 3 models are used in practice. This idea can surely be exploited in some other applications, but it appears to have a limited demand in industry and it has not been pursued much lately. It is the switching between different models in real-time and the selection of the *most suitable* model that makes this study challenging. The effect that different facial expressions and aging factors have on statistical model was another intriguing aspect that was mainly pursued by Lanitis *et al.*[20, 21] who now has another implementation of appearance models [22].

2.9. **Flaws.** Active appearance models are still not as powerful as active shape models in the sense that they require more time to reach good convergence. Furthermore, the accuracy of appearance models is usually lower[19]. However, if synthesis of photo-realistic images is a pre-requisite of the model to be used, then AAM's are a unique and exciting technology that does the job adequately.

It is yet hard to ignore the fact that results of an AAM are less accurate than those of an ASM. This brings up the doubts as for whether the extra complexity associated with texture is worthy of being used. The investment of time and intensive effort, including the need for human intervention raises some important issues.

A significant drawback that is associated with appearance models is that automation of model construction, landmark selection [23, 24] or more fundamentally image correspondence [25] is somewhat difficult. It is not obvious how to choose landmarks sensibly and how to judge the optimality of an automatic choice of points of significance. Since the efficiency of an appearance model depends greatly on the textures embedded in that model, it is not sufficient to use existing techniques to select landmarks and pseudo-landmarks (additional points between the original anatomical or mathematical landmarks) , as quite recently suggested by Davies *et al.*[7]. A further explanation of this work is spread throughout the following few sections.

[19] Although results approve this claim, it is quite likely that better implementations and further improvements will suggest otherwise.

3. Non-rigid Registration

3.1. **Image Registration.** Image registration has become essential in several domains where reliable acquisition of aligned images cannot be assured [26] or turns out to be complex. Needless to mention, the significance of this problem is made most apparent when alignment of a *group* of images must be guaranteed and the images are rather different in nature although they describe the exact same thing.

Misalignment can be due to movement of the object or objects of interest, change of view-point, change in general conditions at the scene and even morphic changes. Morphic changes can be observed over time in the scene or within its constituent parts, e.g. the inevitable change in the form of the lungs. In some circumstances, as later discussed, misalignment is due to profound changes in the form of objects (usually *subject*s) being scanned. Alignment is a key step that must be completed before analysis of the joining of images is safely embarked on. It allows better understanding of the contents of all images as a group – something which is hard to achieve in the lack of *consistency*.

3.2. **Transformations.** Image registration ordinarily involves the manipulation of image pixels according to some rules and under the imposition of several strict constraints. It is usually desirable to obtain a maximum similarity measure amongst a group of images with a minimal extent of distortion. Even a small level of distortion may induce wrong assumptions or violate some stern conditions which should otherwise be an unbreakable prerequisite. It is possible to think of the transformations used as if they pertain to different levels of "interference" – the interference to the analysis and interference with the integrity of the data. A typical categorisation of transformation types is as follows (ordered by increasing interference)[20]:

Rigid: Allows scaling (uniformal size changes, i.e. shrinkage and enlargement), translation (location in space) and rotation. The normalised shape attributes are altogether preserved and the process is usually concerned merely with some common alignment and bias-neutralisation. Such alignment usually aims to place all instances

[20]The names infrequently change in the literature despite standardisation. What is important is the description of transformations and not the names or mnemonics that wound up describing them.

upright and centred in the space origin with a fixed size of maximum 1 unit. The instance is virtually confined to lie inside a bounding structure (a square, or sphere in 2-D and 3-D respectively)[21]. In 3-D, for instance, there is a total of 6 degrees of freedom so a rigid transformation will be wholly characterised by a tuple of 6 values[22].

Affine: Allows the instance (image) to *stretch* along at least one axis or dimension, but not necessarily all (so that homogeneous scaling can be broken). Despite the fact that previously essential constraints are broken, all lines that were parallel remain parallel after the transformation is applied[23]. Reconstruction is said to be possible so that this transformation is invertible. For all affine transformations $T_a(x)$ where x can be a vector representation of an image (or volume) and their inverse $T_a(x)^{-1}$, the expression $T(T_a(x))^{-1} = Id(x)$ must hold. This relation must always be calculable and retain simplicity which makes it easy to resolve. This will prove to be an important constraint when the practicability of warps is debated.

Non-rigid: All other valid transformations fall into this category. This includes tapering, spiral warps, pinching, etc. In principle, no inviolable constraints are in place, but quite clearly the transformation attempts to preserve some of the primary structure[24] of the image while avoiding tearing and folding [27]. This means that each pixel in the range must map to another and no pixel is left undefined. A bit more on this is to be explained later.

The image below illustrates the effect that each transformation is allowed to have on the image on the left.

[21] Such a process is a very fundamental one in computer graphics modelling and various books cover shape-normalisation techniques and algorithms.

[22] 1 for scaling, 3 for *x*, *y* and *z* coordinates and 2 for rotation, e.g. the *xy* and *yz* angles θ_1 and θ_2.

[23] Popular transformations such as skew, shear and taper, on the contrary, are not parallelism-preserving. The importance of this rigorous constraint is that the distance between any two points remains proportional to the transformation.

[24] A random uncontrollable transformation will dispart basic structures in the image and make interpretation impossible.

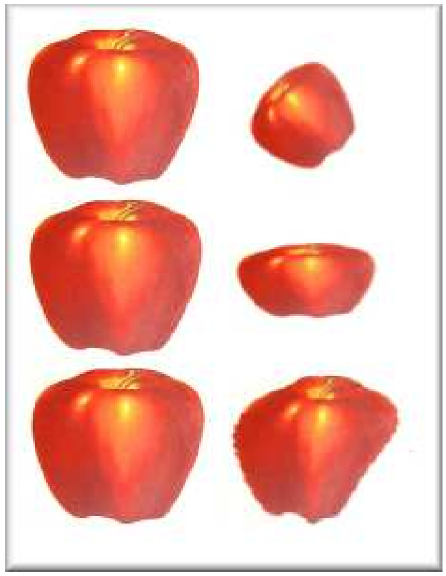


Figure 5: Registration examples; from top to bottom: rigid, affine and non-rigid registration.

As the figure suggests, the appearance of an object remains identical under rigid transformations. It is allowed strictly to grow, shrink, move, and rotate. Affine transformation allows an object to lose its original form and non-rigid registration is far more permissive so the object can be subjected to rather obscure deformations.

What follows in this section briefly explains some of the main concepts, techniques and ideas currently devised. The actual key points, which describe non-rigid registration in the context of the Structure and Function Grand Challenge[25], are as follows:

(1) Warps
(2) Similarity
(3) Objective function

These three points will be explained in a bit more detail with reference to current work, practical considerations and attempts made. For now, a concise introduction would do. The approach often taken is that an image need to be warped (equivalent to transformation) until it matches another. The match is estimated with the assistance of similarity measures and this

[25]The Grand Challenge aims to unify the different stages of analysis. It will be referred to yet again in Section 5 which deals with recent and on-going work.

process of warping and similarity is sometimes wrapped up and put under one generic objective function. The objective function is then handled by a general optimiser – a term which is further explained in Subsection 3.7.

3.3. **Diffeomorphism.** The concepts and arguments introduced in this section show why there is an ever-increasing interest in non-rigid registration, based on non-rigid transformations[26]. The mathematics behind the required transformations and the theory that needs to be established in order to make them practical is constantly being explored and papers on the subject receive attention and recognition. Diffeomorphic functions are *invertible, continuous* and *one-to-one* mappings for a given image[27]. These mappings are usually described by some local geometrical transformations that have an effect on pixels or the plane that pixels are embedded in.

Current transformations that are used in Manchester University by Twining and Marsland [28] also benefit from having continuous derivatives at the boundaries, unlike for example, Lötjönen and Mäkelä [29] who suggested a similar transformation. This, however, is a convenient property that is not a necessity. It is just a strategically good attribute to have in real-world applications.

What invertibility, continuity and one-to-one mappings mean in simpler terms is that for each transformation:

(1) That transformation has an inverse so that any transformation (or *warp* as it will be later referred to as) can be reversed.
(2) That transformation affects *all* data (pixels) within its boundaries so it has a centralised yet somewhat global effect[28]. This means that every point must move as would be expected to give a continuous flow of intensities. The transformation should also be cautious not to corrupt structures in the image in any way.
(3) No two points should be mapped onto the same point as this would "strip off" areas in the image.

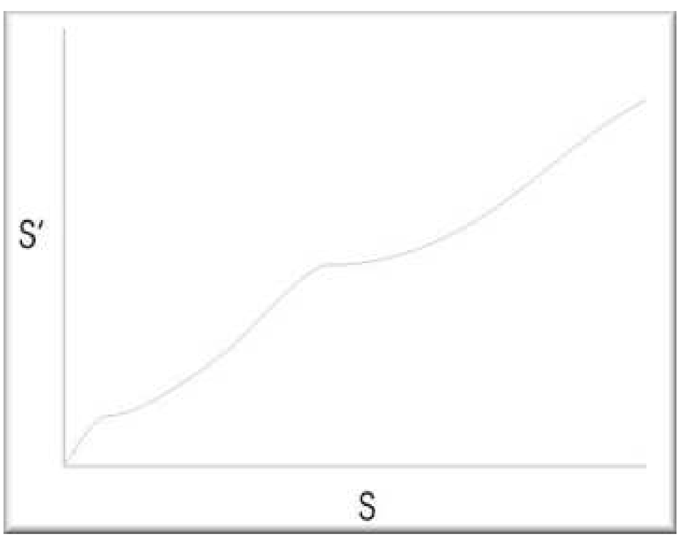
[26]This so-called mapping or transformation can be thought of as being a standard function, for example $f(x, y) = (x', y')$ in 2-D and it is applied to all the pixels within a predefined range.

[27]More generally, the functions are mappings defined over a matrix or a vector which is analogous to an image.

[28]A pixel of course can be mapped onto the exact same original position, but the idea is that a continuous flow must prevail.

3.4. **Reparameterisation.** A shape can be described by a collection of landmarks as shown in Figure 1 earlier in this report. The landmarks are usually located at corners, T-junctions and edges that are easy to locate and other points in between these landmarks are chosen to expand the representation of that shape and make it richer. To register multiple images, all corresponding landmarks and points must overlap in as accurate a way as possible. They must correspond to one another on one common grid so that image analysis can proceed. One way of doing this is to apply diffeomorphic warps to the space in which images will be embedded. That newly-defined plane is supposed to bring the collection of landmarks across the set of input images closer together. This ends up bringing a number of images to correspondence of some quality.

It is not, however, obvious what choice of landmarks and intermediate points will finally result in an optimal overlap or even a good one. To automatically shift points and evaluate the the subsequent global (or pair-wise) effects, *reparameterisation* of these points must take place in a way that preserves their order along the imaginary contour they form. A new spread of the points needs to be chosen iteratively and the results recorded. The spread of the points can be defined purely by a function and the reparameterisation alters this function to find preferable results. A monotonically-increasing function describes the distance of all points[29] from an arbitrary point on the curve in such a way that will not violate their sequential order.



[29] A continuous function is independent of the number of points. Therefore, the complexity can be increased progressively to obtain finer, more accurate results.

Figure 6: Monotonically-increasing function

Figure 6 shows what is meant by a monotonically-increasing function. The following must hold although its inverse may hold instead (a monotonically-decreasing function):

$$\forall(u \in S \cap v \in S \cap u < v) \rightarrow f_{mon}(u) < f_{mon}(v) \tag{3.1}$$

where f_{mon} is the monotonically-increasing function used and $f_{mon}(S) = S'$. More simply, the derivative at any point must be positive, i.e. $0 < \theta < 90$ so that $0 < tan(\theta) < 1$. In the figure below, the meaning of reparameterisation as it is applied to points of a shape is made clearer. The distance or offset along the curve is guided by the value which was determined by the function above. In this particular way, all points which lie on the curve can be moved *simultaneously* without colliding with one another and new autonomous descriptors of shape become available. Instead of describing the movement of each individual point, an arbitrary number of points can be shifted according to one modifiable function. Davies *et al.* used this technique to optimise a shape model by appraising any selection of points. For each such reparameterisation, the specificity, generalisability and compactness were evaluated. The first and second of these terms were coined in the thesis published by Davies.

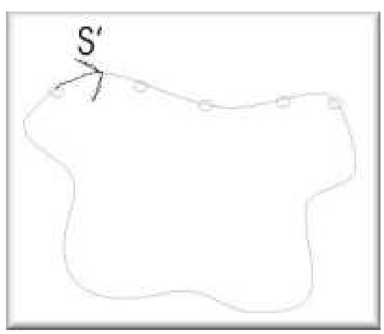


Figure 7: Reparameterisation example

Prior to the invention of this technique, points were often chosen to be put on the curve so that they are equally spaced. This approach was often a straight-forward and computationally inexpensive, but its results were unsatisfactory for more complex shape where the curve bends sharply. Some attempts were made at placing more points at regions of high-curvature, but these were still inferior to the aforementioned reparameterisation-based approach.

3.5. **Warps.** This short Subsection adheres to a local perspective – a perspective along the lines of which future research should move. Subsection 3.3 on diffeomorphism introduced functions that map a group of pixels to new positions. These functions will now be referred to as *warps* plainly because this is the terminology that is usually used in the literature. Due to practical considerations, the warps used are chosen to be rather simple and therefore computationally inexpensive. Some will argue that more sophisticated warps will produce better results in a smaller period of time because a smaller number of these is required to reach overlap as explained previously. However, they may also damage some structures in regions that are better left untouched, as well as interfere with previous warps that supposedly did the right thing.

The warps currently used are round (and extend to spherical) and they can be described by their location, radius and possibly depth. Many such warps are applied at different scales to the image. Their position is quite random and good results are committed and carried on to later iterations while bad ones get discarded. Towards the later stages of the algorithm, only small local warps, much as in the case of reparameterisation, will give the most qualitative results.

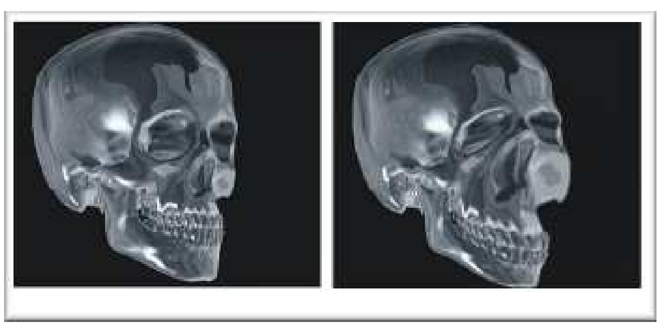


Figure 8: Warp applied to image. On the left: image before warp is applied; On the right: image after warping.

The choice of warps is quite arbitrary and their most preferable complexity level is still an issue of active discussion[30]. The randomisation of the

[30]In Manchester University, Cootes and others are in favour of many small warps, but some are in favour of few rather more complex warps that are controlled by a larger number of parameters. More details on such issues appear in later discussions on current work.

location at which warps are set means that computation is saved on understanding the image and using that priorly-gained understanding. On the other hand, many warps are discarded in this way and wasted effort makes this method far less elegant.

3.6. **Measuring Similarity.** There are various ways of measuring the conjectured similarity between two images. Mean-squared-differences or sum-of-squared-differences are rather poor methods of getting a useful measure of similarity if the images are far away. Therefore, such measures are often better off used when convergence is foreseen. There are also measure that are immune to large spatial displacements or variability in form. Histograms of the intensity values in the images, where intensity values are accounted for globally (or locally, inside regions that require greater emphasis), would be far better measures under most circumstances. Extra strategic steps, such as the removal of empty bins in such histograms, are taken to make the histograms more powerful indicators of similarity. Active research repeatedly reveals better algorithms as will be described in brevity below.

Although the correlation ratio is still occasionally used to measure similarity, it is less relevant to this report and goes back over half a century ago [30]. Mutual information and normalised mutual information, as described by Studholme, give good measures that see high usage in existing non-rigid registration algorithms. Each one will be dealt with in turn.

3.6.1. *Mutual Information (MI).* Viola [31] has developed a way[31] of finding and measuring the similarity between two images or more by repeatedly comparing couples of images. This ability to compare images is crucial for registration of images as it robustly and accurately returns an estimate of the beneficialness of the warps applied. Section 4 deals with the introduction of information theory and some of the basic measures which can explain the following in more detail. However, it is the principle that is worth understanding at this stage, rather than tedious technicality [32, 33, 34].

Mutual information computes *volumes of overlap* in images. If two images are matched, the joint histogram is then expected to give an indication of where *sharp* grey-value peaks are located and the sharpness value of these peaks. Under the converse case which is mis-registration, the joint histogram is then expected to show peaks of low sharpness and new peaks can emerge. The algorithms and advanced information theoretic expressions

[31] The discovery of this mutual information is actually attributed to Maes as well. The thesis worked on by Viola in the mid-nineties received great recognition though.

that take advantage of this observation are at this stage left out entirely. At this point, it is only worth defining a joint information (or entropy) to be $H(A, B)$ and state that MI calculates $H(A) + H(B) - H(A, B)$. This means that mutual information is subtracted from the sum of information present in the two individual images.

3.6.2. *Normalised Mutual Information (NMI).* Studholme [35] and Maes[36] suggested that some normalisation should be applied to mutual information as was described above. Quite a few steps are involved in this normalisation process and the full mathematical summary is left to the literature[32].

The main difference is that the expression used for MI is significantly extended and divided by a normalisation term. The method is predominant in non-rigid registration as it yields good results

3.6.3. *Mean Sum of Differences (MSD).* This measure was explained and illustrated in the context of active appearance models where difference needs to guide model fitting. Its idea is primitive, nevertheless it is effective, especially when faced with the simplest class of tasks. Pixels are compared in two images one by one, their squared grey-level difference is calculated and a sum[33] over all differences is returned. This method is usually powerful if the two images compared are closely aligned and their intensity values are relatively continuous and low in contrast. In other words, MSD will tolerate only a low level of locally-situated difference, while contrariwise, MI and NMI rely on sparse dispersion of all pixels.

3.6.4. *Acceleration of Similarity Measures.* As earlier described in the context of active-appearance models, a multi-resolution approach is used to speed up the whole process. Blurring or averaging followed by scaling allows for images of smaller size to be manipulated and complexity to be quadratically reduced. As the similarity measures are proportional to the images size, far better performance can be achieved by a transition from coarse to finer resolution. Pluim [37] identifies the effects that this approach will have on the measurement of similarity.

[32]There is an additional distinction between symmetric and asymmetric normalised mutual information, but explanation on this requires the full technical recipe. The dissertation at `http://www.lans.ece.utexas.edu/~strehl/diss/node107.html` summarises the way in which NMI evaluated.

[33]One could suggest an extension to such a method and assign weights to differentiate regions of varying significance.

3.7. **Optimisation.**

3.7.1. *Background.* General optimisation is often used in the process of matching and its complexity can be relatively high. The behaviour of such a problem is not linear and it may cross over to the realms of quadratic programming (QP) where various parameters simultaneously control a function and minimisation is therefore by no means trivial. This process is by convention concerned with the minimisation (complement is used to generalise to maximisation) of the value of a function and that function often comprises more than a single variables which makes it multi-dimensional. Many software products that act as general optimisers exist and the way they operate and perform varies. Some even switch between different algorithms depending on the stage of the optimisation and the changing granularity of the problem.

Optimisation over a function which varies in many dimensions is an expensive process. Often this optimisation shall require some *a priori* knowledge of the problem domain so that performance winds up being satisfactory. In the case of image matching, advantages can be gained if the effect of variable alteration can be predicted in some way. An example of this was described in Subsection 2.5 where pixel intensities have a dependency upon a group of parameters. Slightly less specifically, given the difference between two or more images, or even some generic data regarding a *change* caused by value changes in the function considered for optimisation, it should then be possible to determine paths that lead to quick convergence.

For the problems outlined in the document, common optimisation methods are gradient-descent and downhill simplex. However, many other methods exist[34] and whole books have been written on the subject [38]. The advocated strategy would sometimes be a utilisation of mixtures of different methods with rational choice of the most relevant one at each stage. That is plainly because the different characteristics of the methods make them advantageous at different states throughout the entire optimisation process.

3.7.2. *Problems.* One of the main flaws of existing optimisation methods is their inability to find a global minimum (or minima) fairly quickly without some additional knowledge about the function under investigation. Rough assumptions about the behaviour of the curve along each of the axes are otherwise made.

[34]To name several more methods: dynamic programming, genetic algorithms, Powell's, simulated annealing and steepest descent.

The pace of the optimisation process can be boosted on the expense of overall accuracy and error likelihood. Sometimes these cannot be jeopardised, mammography being an example of choice. It turns out that if no exhaustive search[35] is carried out, there is then a danger of convergence at some local minimum. In most applications, any stoppage at a local minimum would be highly undesirable although this may be better than a complete failure at identifying low points. Local minima are a necessary evil for large and complex continuous functions.

In conclusion, there is a trade-off between speed and accuracy although accuracy can be achieved at a lower cost if more knowledge is acquired offline, before the optimisation task actually begins. Quite expectedly, this also implies that many redundant computations will consume precious resources and time in order to train the optimiser.

3.8. **Objective Function.** The objective function is the actual function which needs to be minimised in order for an optimal choice or a solution to be picked from the many alternatives offered. The function is most heavily based on similarity measures as briefly explained earlier, but it allows this measure to be extended in some way. For example, it can be helpful to include the computational cost of the warps that are used. The reason why the cost of the warp is sometimes an integral part of the function is that long-winded warps are not nearly as desirable as simple ones that perform the task equally well or even better. This cost is often considered a *normalisation term*.

Objective functions are built to encapsulate in a concise and effective way everything that is repeatedly evaluated. They are therefore required to be a very efficient unit which will be invoked quite frequently. The *speed* of the registration will directly depend on the choice of an objective function that adds up results from warps, similarity calculations and possibly more components, as can be seen in current group-wise registration papers. The *quality* of the registration will of course depend on this function, too.

Let us define two images I_m and I'_m to be the images before and after warping respectively. Let us also define a warping function $f_w(x)$ to be $f_w(I_m, < parameters >) = I'_m$. For a similarity[36] function f_{sim}, the objective function can then take the form:

[35] Exhaustivity is impossible for continuous functions, but digital images are luckily discrete.

[36] It will temporarily be assumed that for an objective function that needs to be minimised, the similarity measure will return small values for good similarity and vice versa.

$$f_{objective} = f_{sim}(f_w(I_m, < params >), I'_m) + < norm - terms > . \tag{3.2}$$

The function then attempts to find the parameter values that will lead it to a globally minimal solution. More precisely, it attempts to find *assignments* for all parameters that describe the warps so that similarity is maximised.

The explanation on the objective function concludes the algorithmic approach that registration takes. Non-rigid registration algorithms can be assessed by methods such as the one described by Warfield [39].

3.9. **On-going Research.** Rueckert *et al.* [40] describe statistical deformation models (SDM's) which are in essence surprisingly similar to active appearance models. Much work has concentrated on using the knowledge and techniques from each one of these two to establish a more powerful framework of full appearance statistical models. The work is described in Section 5 with reference to research that is associated with the GC. An exclusive introduction to Rueckert's work will shed some light over the current registration concepts which future research relies upon.

Non-rigid registration methods have been applied in several medical domains of expertise. Among these is the renowned brain analysis task, contrast-enhanced MR mammography and segmentation and tracking of the heart. The procedures currently employed are inclined to follow higher-order entropy measures that will not be delved any further. Rueckert's homepage which is listed in Appendix A gives the full details and references. The next section on information theory explains in brevity some of the basic ideas behind these so-called entropy measures.

The success of temporal non-rigid image registration method is dependent upon two factors:

(1) **Search algorithm:** As earlier illustrated in the context of active appearance mode, good warps need to be searched to achieve good similarity.
(2) **Similarity:** The performance relies on a suitable choice of similarity measures which guide the search until a sufficiently good fit is declared.

Learning the properties of similarity measures, the way they affects the search duration and the effect warps have on similarity are all important aspects of registration method development. This is reminiscent of the process in which correlation between parameter changes and intensity changes are learned in appearance models.

Change in organs due to resection (craniotomy being a frequently-encountered scenario), expansion, movement etc. is often modelled using thin-plate splines and the motion of organs can be handled using free-form deformation (FFD) which are based on B-splines. Prior to this embedment of high-order functions, the effects of rigid-body motion is annulled by Euclidean transformation. Similarity measures guide this process of rigid registration just as well. It is the technical description of the algorithms used that proves why these methods, which are used in Guy's Hospital, are extremely effective. Future reports will address the finer technical details that express the operation of the algorithms.

4. Information Theory

4.1. **Importance.** The arguments regrading the importance of information theory with respect to this project vary. Information theory is indeed valuable due to its relevance to past projects in the field, on which future projects will rely. Image analysis is often involves the passage and handling of large sets of data and extraction of the meaning of the data is a necessity. Compression becomes ever more crucial when large models and entities are maintained in memory and, again, reasoning about compression goes back to the theory of information. New ways to encode data, avoid redundancy and describe objects succinctly are being sought as they often reduce the *complexity* of any system as well as its *size*. Measures of information are necessary to introduce and support learning capabilities which in turn form intelligent systems. Such systems can evaluate and judge improvement as illustrated thus far and as will be illustrated later.

4.2. **Shannon's Entropy.** The term entropy is used to denote a general measure of *uncertainty*. It is not a very sophisticated idea, yet a very fundamental one which was first introduced in 1948. Uncertainty is associated with the required amount of data so it can also be thought of as an *information measure* or quantifier. The value that quantifies uncertainty originally related to random variables which take different probabilities among a set of states (reminiscent of Markov chain models). Shannon's entropy has become a very useful way of evaluating structures and pattern in some data. The lower the entropy value, the more data is already inherent in that data. In a sense, the entropy indicates how much can be learned from the data and what is still unknown.

4.3. **Minimum Description Length (MDL).** Minimum description length [41] provides a measure of the *minimal* amount of information necessary to encode some data. Any data can be transformed in a particular way so that it becomes a sequence of symbols (numbers or signals even, to be less general)[37]. The transition between one symbol to another can be encoded by some transition table which holds the probabilities of all possible transitions. For n symbols, up to n^2 transition need be defined. Markov chains are one such model type which is a convenient way of explaining the nature of MDL. Markov chains of a high order can accommodate for data of more

[37]Binary representation is quite complete in the sense that any data, e.g. programs and text, can be coded in a binary form. However, this representation might be very greedy of space and the issue of representation compactness then arises.

awkward and unpredictable variance. MDL infrequently defaults to higher-order models which are superiorly expressive, although they require a far greater number of transitions to be specified.

With proper use of models as in the case above, data can and should be represented solely by all transitions and can then essentially replicated from these transitions. Unless the data is peculiar and shows no patterns, such a description would be compact for data large enough in bulk. MDL attempts to describe the extent to which some data is capable of diminishing in bulk (with or without loss being a separate issue) or rather the minimal amount of information that needs to be available to describe and reconstruct that data. In most cases, the more uncertainty present (i.e. higher entropy), the greater the minimal description length would be.

As an example, here is a vector representation of some arbitrary data: $v = \{3, 4, 3, 1, 1, 1, 3, 4, 2, 2\}$. There is an alternative way of representing this data. By using a first-order transition table, e.g. $3 \rightarrow 4$:1, $3 \rightarrow 2 : 0$..., the likelihood or probability of transition from every element to its successor is revealed. Observable patterns are merely meaningless in this data example. Encoding of transitions will also be an inefficient approach as a result of the small vector size and the low sequential correlation[38]. On the sharp contrary, vectors such as $v = \{0\}$ or $v = \{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}$ bear a very small measure of uncertainty. In the second of these[39], only one transition exists so it can be represented by a tiny model and the entropy is 0.

To summarise, MDL is a measure of the minimal amount of information that expresses a sequence[40]. By inspecting transitions it is possible to get an insight into the complexity of some model. A heart beat pattern, for instance, is rather predictable and repetitive in comparison with the positions of a person's fingers over time. This means that the description length of the heart state should be shorter than that of the hand. Less information is required to capture the behaviour of the heart in motion.

4.4. **Minimum Description Length in Modelling.** The concept above has been applied to select good descriptors of shape [42]. Selection of points

[38] To make this appear more practical, one can think of a large ($>$ 100000 pixels) image where patterns are present.

[39] This can be portrayed as a uniform plain-white image.

[40] General problem reducibility to a sequence is axiomatic as Turing Machines suggest.

that describe a given shape, as explained in Section 2, was perpetually altered and evaluated to find shape models and examples that require a smaller set of data to be passed as an encoded message.

To express the process at a moderate pace, each time points on the curve that trace the shape are selected, a different model is ultimately constructed. A good and compact statistical model is one whose legal variations are relatively small and possibly so are the number of its control points. Such a model is sought via a general optimisation regime under which point are reparameterised. MDL can be used in an objective function that is iteratively evaluated for each such points reparameterisation. The minimisation process was described in reasonable detail in Subsection 3.7 on optimisation. The more genuine part of this work is the use of an existing information theoretic measure, namely MDL, to guide an autonomous search for good models. This work will be explained with respect to current research in the next section.

5. ACTIVE APPEARANCE MODELS AND NON-RIGID REGISTRATION

5.1. **Existing Work.**

5.1.1. *Synopsis.* The work of Katherine Smith investigated warps on a simple bump-like curve (see Figure 9 below).

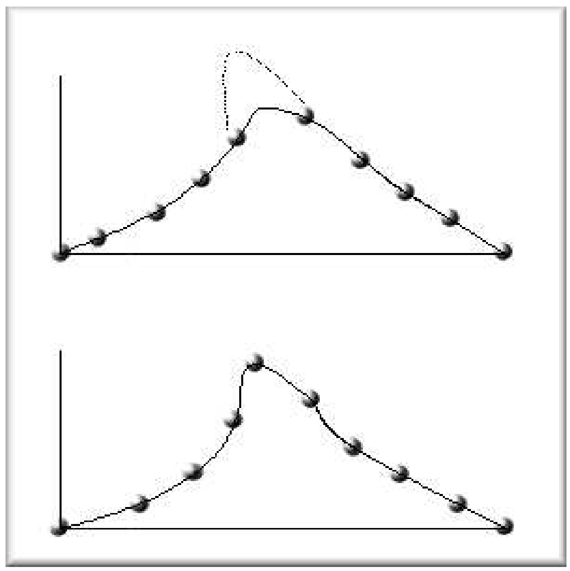


Figure 9: Reparameterisation along the curve

Description: The first step taken by the application was the generation of some random bumps simpler than the one above. These bumps varied in height and width although the property of height was intended to be ignored during registration. The bumps were all symmetric and the height is one of {*hi*, *low*} where $lo = 0$ and $0.7 < hi < 1$. The data was therefore far simpler than any 1-D data which is not constrained in any way. The height of the bump and the position at which the bump goes high can conjointly define that bump so two real numbers at the minimum would suffice to derive each bump.

As images are being warped, the form of the bump quickly changes to give a smoother curve with more continuous derivatives. This of course will depends on the type of warp which is applied to the bump. At each iteration, new similarity with respect to some reference image or similarity with reference to the whole set of images is obtained using warps and measured using the methods outlined in Subsection 3.6.

The similarity measures used in these experiments to evaluate similarity were mean-squared-difference (MSD) and mutual information (MI). The latter was more computationally expensive so although it gave better results, it needed to be used with caution. Likewise, the type of warp applied was often but not always a simple one which is controlled by a single allocated control point. In some cases, many control points were assembled to form an expensive warp of increased complexity. The choice of these points was often decided to be random as a successful rational choice would have required much more speed, consequently slowing down the whole process.

As explained at some capacity beforehand in Subsection 3.4, reparameterisation was used to perform points placements in the image of the bump. These points did not directly express the form of the bump, but rather controlled the *warps* that affected the bump point coordinates. Initially, the curve to be reparameterised was an ordinary linear function stretching from the origin to a point (n, n) where n is the number that is chosen to be the image width (the only dimension of the single-dimensional data). Points were later chosen according to the change imposed on the curve due to warping.

The experimentation Smith carried out allowed for many combinations of different options to be set, applied and appraised comparatively. The estimates of the "goodness" of warps were evaluated using the creation of an appearance model from the group of images at present state, making this a group-wise optimisation methodology. This was not the first time that such an approach was investigated as Subsection 4.4 shows.

The images after warping had been applied were treated as training data for the creation of an appearance model. PCA reduced the complexity of that model as required. The compactness of the model which could be derived from the the sum of variances or the determinant of the covariance matrix[41] was then scoring the choice of warps after they had been applied. In this way, a better choice of warps could be made so that bad ones quickly get discarded and the state of all affected images reverted.

Summary: As the above descriptions imply, this work was able to show how statistical models go hand-by-hand with non-rigid registration. In this case, they simple *evaluated* the (non-rigid) registration process and distinguished between the many alternatives offered by different families of warps, similarity measures and so forth. Needless to say, the run-time became a real difficulty when ill-chosen strategies were attempted. Smith took

[41]This will indicate the *volume* of the model's scatter in space. The more compact a model appears, the lower this volume.

this into consideration in the final evaluation and comparison of all different experiments.

Advancements: The work of Marsland, Twining and Taylor [43] went a step ahead and investigated a full 2-D model. However, it concentrated on just a simple contour (defined by 12 control points) of the skull shape as pictured from an overhead perspective. The figure below shows that warps can have an effect on the *whole* shape, but still lack some control over local structures such as the ventricles. Varying scale can solve problems like this and make the global non-rigid registration approach very robust. While this work produced elegant results, it did not explore many varying options as Smith did.

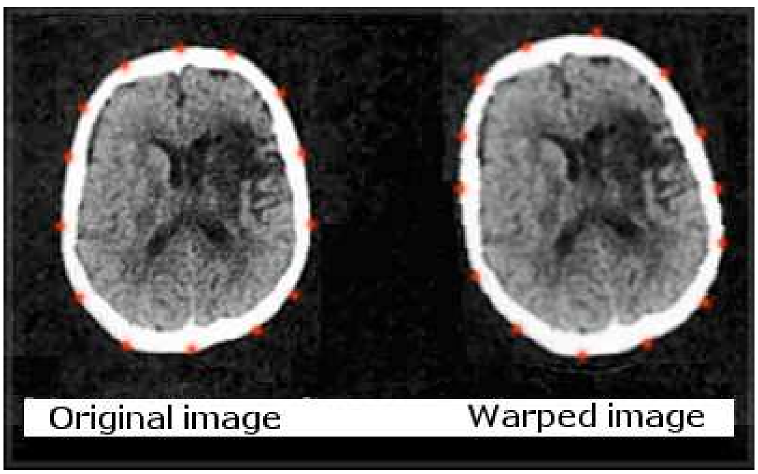


Figure 10: Brain image warping

5.1.2. *Drawbacks.* Drawbacks and gaps do exist and controversial implementation decisions are listed later in this section. There is much work to be done in order to find out which the better performing approaches are and the experiments and applications described above provide a substantial starting-point.

5.1.3. *Alternatives.* The main alternatives to non-rigid registration are rigid and affine equivalents, but these are rather impractical. In most real-world applications such as registration of brain-slice images, there is a very slim chance of getting satisfactory alignment of structures while preserving some continuity unless non-rigid transformations are applied. One may argue that affine registration should suffice, but what if parts of the brain expand beyond proportion? It therefore appears as if, from a registration point-of-view, no obvious alternatives are yet known. The ones mentioned above give the best performance yet and comparison with the closely-related active appearance models suggests that flexible deformation is mandatory, especially for biomedical data.

5.2. **Previous work.**

5.2.1. *Summary.*

Non-rigid registration (NRR) and model-based image analysis were previously believed to possess some commonality – a premise on which current research is still based. There is a growing belief that the best of both can be exploited to construct a unified framework. This modified framework might be more robust and offer higher utility and functionality when compared with the other two approaches working solely.

It is claimed that warping of the images, as is already done in medical imaging in particular, can be used to find correspondences that are optimal in some respects. Group-wise image registration using non-rigid transformations was the way in which previous research attempted to build good models of appearance. Furthermore, non-rigid registration made it possible to achieve better correspondence in images and appearance models could highlight desirable registration, i.e. ideal warping sequences. As a result, better active appearance models could be constructed and non-rigid registration was guided by appearance models rather than similarity measures.

5.2.2. *Results.*

The results of previous work that dealt with group-wise registration and globally optimal models were clearly better than those obtained from pair-wise registration under similar conditions. This comes with a cost however. The improved process was far slower and the results were not superior by a very noticeable measure. In fact, it is quite possible that the results were partially biased due to the conditions set for the experiments.

Impending experiments aim to disprove the claim that the conditions and choices made were the cause for the apparent advantage. They just as well aspire to increase the apparent gain of group-wise registration as it is not yet evident. Various improvements and lines of research have already been suggested and similar work continues even at present.

5.3. **On-going work.** Work in the field seems to move in differentiable yet almost identical directions. On the one hand, speed is an issue that might not have a solution and Section 7 elaborates on imaginable hindrances of this kind. Algorithmic trade-offs and different choices of programming language and paradigm, operating system and platform is a matter worth pursuing. Since the process relies on wide global scope, i.e. investigation of various images simultaneously where change in one affects all, heuristics can perhaps be applied to decrease the number of iterations involved[42].

Claims of a similar nature can be made on the more intrinsic part of work being reviewed. For a start, values were often tweaked manually and no strong evidence was used to support such arbitrary selections which purportedly followed common sense. Another problem that has been realised is that much of the process comprised the simplistic joining of remotely-related components whose nature is unique and autonomous. This means that components in the system often suffered from the undependable fusion which was in place – something which is a direct consequence of the knowledge that is still lacking in the field. There is much to be learned about how the numerous existing techniques, measurements and component should be merged effectively and by all means conveniently.

5.4. **Alternatives.** There are more than a single alternative to be looked at and many different aspects call for attention as the earlier parts explain. Here is a short summary that may help guide future endeavours:

(1) **Speed-up:** The methods operate very slowly for most globally-driven approaches. A solution to this is desirable because not only would it stimulate more experiments and experiment feedback, but it would also make these methods usable and marketable.
(2) **Data extension:** The simple existing bump which is generated in MATLAB needs to be extended, possibly by conversion to a smoother bump as the one described in the research of Davies and Taylor.
(3) **Lambda:** Lambda coefficient in the objective function weighs appearance against shape and its value is subjective to the problem. Experiments can find alternative solutions or better assignments for lambda.
(4) **Automation:** It would be desirable to create a (compilable) system that copes with the full cycle of analysis without outside intervention and without any pre-existent data annotation.

[42]Although knowledge of the problem is an integral part of most program optimisation steps, the more formal methods can be used to identify dependencies. A dependency graph can reliably indicate where re-evaluation is indeed necessary.

(5) **Generalisation:** Many *ad-hoc* algorithms are currently used for group-wise registration. An more impressive system would deal with arbitrary data without compromise to the quality of the results.

5.5. **Relevance.** Along the lines of past research, the possible developments and gains all appear to offer various advantages to this system of registration and analysis. Unlike many other such projects, this one is open-ended and is dependent on the outcomes, discussions and discoveries published in conjunction to one another. As a group effort is expected, a high level of interaction and communication will be involved. This has clearly been the case to date.

5.6. **Significance.** If this research reaches and obtains its goals, more theoretical and practical grounds will be available for future applications of non-rigid registration in active appearance models and vice versa. Such goals are quite important as appearance models have not yet become as wide-spread as one could argue they should be. By making real use of the intensity information that is available in appearance models, an exceptional practical strength is exploited. Full image synthesis is an application where no other model type can yet be seen as a substitute. With more automation in place, higher accuracy, compatibility with other technologies, awareness and ultimately, ubiquitous use of the technique can be shortly anticipated.

It is worth pointing out that availability of active appearance models to individuals who deal with registration of images would take this technology one step ahead. This will introduce more concepts, metrics and studies which increase their functionality and flexibility. Reciprocally, previously "foreign" techniques can extend the functionality of active appearance (and maybe shape) models once they are put in the hands of groups with difference background and expertise.

As an instance for the first of the two contributions above, a radiologist could very comfortably view a highlighted model of an organ that is deformed in a natural and sound manner. Results and analysis can be managed rather quickly and neatly as automation and synthesis generation should be made operable and even interactive[43].

[43] A reasonable response time depends on the purpose of the system, the level of detail, etc.

6. Potential Developments and Goals

6.1. **Starting Point.** It is worth starting off with a description of the most recent and most relevant developments. The following few paragraphs summarise and highlight some of the main principles which describe the methods used in the existing systems. These are the systems that surely need to be extended and their understanding is the most crucial.

Smith's work follows the work of Davies in an obvious sense, but it explores a different domain with slightly different aims. Davies repeatedly performed a reparameterisation over a given series of shapes, or rather their defining points. All these points were shifted in accordance with displacements, as orchestrated by a monotonically increasing curve. This reparameterisation was applied to all examples in the training set to evaluate an optimal choice of point spreads, or more precisely, the favourable reparameterisations that act upon these points.

In current group-wise registration work, the elements that such reparameterisation affects are the points which control the *warps* applied to the data[44]. These chosen warps are then applied to all the examples (or data instances) and measures are used to describe how "similar" the data is *collectively*. Different measures of similarity are used as well as different types of warps. Another way of explaining this process is to say that warps are being found that make data lie in similar positions in the imaginary image grid. A warp implicitly defines an uneven plane for images to be embedded in and when all images get embedded in that plane, they should then be collectively similar. Interestingly, that similarity can be checked with the use of AAM's. Ways of evaluating an appearance model and ways of drawing conclusions about the data that was used to build it already exist. The algorithms developed for this work use a similarity measure such as MSD or MI to see how similar images become during search[45], before a model is created. The model created from *all* the examples is the entity that defines the "goodness" of the warps. A model can in some sense describe and measure of similarity across the entire set, as oppose to the pair-wise measures using beforehand. This construction of a model can in this way guide the search for good warps. MI and MSD are utilised to check local, small-scale changes only. The system seeks control points that define good warps and it seeks such points using the idea of reparameterisation. The

[44]The data type is irrelevant. It makes no difference whether it is an image of full appearance or just a bump.

[45]This similarity computation is incorporated in the objective function and it usually comprises a collection of pair-wise similarity measures.

resulting warps must then produce good appearance models for the *whole* data. For example, in the case of these specific experiments, all the bumps are warped to become quite similar so the model created from them has a low determinant.

6.2. **Goals.** A main goal, which relates to the big picture that is the GC, is the merger which involves (non-rigid) registration and statistical models. In both cases, some *dense correspondence* across some or all of the images is involved and must eventually be determined. Reuse of the information that is incorporated in each of this two techniques (which are believed to be inherently the same) would make the overall analysis task more powerful, flexible and well-integrated. If even a moderate combination of the two is obtained, then new ways of building and using models will be open for investigation. The parallel development in both fields, especially the need to identify homologous structures, is what makes this GC so suitably arranged and increases its potential of resulting in success.

In NRR, lower-level inspection of image pixels identifies similarity using mutual information (or any other similarity measure for this argument's sake), whereas in statistical modelling, the correspondences are often marked by hand (as explained in previous sections this is no longer quite the case) or gathered in an ill-chosen fashion. It is imperative that effort is made to reuse the segmentation from NRR so that models can be constructed more quickly and fitted to targets before feature extraction takes over and does its part of the analysis job.

6.3. **Developments.** The simple data used by Smith is already proving slightly too cumbersome for responsive experimentation on a relatively strong machine (1.8 GHz, 512MB RAM), especially owing to the complex algorithms devised for group-wise registration. It is advisable that evaluation via profiling toolkits is firstly made to hasten this process as much as possible. Alternatively, coding of the algorithm in a compiled language as C++ is seriously looked at as a possibility. The complexity of the departmental VXL library is believed to make such step less than desirable.

Once speed-up has been taken care of or when it is at least known that a near-flawless well-performing piece of software is at disposal and under control, the simple 1-D data can see the addition of a few additional characteristics. This new composite[46] data must retain some good commonality

[46]It will be prematurely assumed that the new *synthetic* data posses several distinct morphological attributes.

and similarity across the set of images and it should not be overly more complex and unpredictable in comparison with a simple bump. A double-humped curve, a round smooth line or even a contour of a a profile of a face might be sensible and more challenging choices[47]. In any case, whichever synthesis of data is eventually selected and experimented with, the choice of control points for the warping then becomes a more crucial issue[48]. A more localised control via warps then turns mandatory because several separate structures exist in the data.

The experiments of Marsland, Twining and Taylor have already shown the realistic application of warps to a medium-resolution two-dimensional data. Nonetheless, it is vital to point out that an elliptical shape was dealt with and *a priori* knowledge of the problem was used to increase the speed of the group-wise registration process. Control points that characterise the warps were initially places on a circle whose centre was the image centre and radius corresponded to the typical position of the skull in standardised imaging. If the problem involved point selection for, let us say, knee cartilage and no knowledge about the object was available in advance, the results would have then taken far more than 10 hours to obtain (as was the case for the 12 points distribution around the skull's exterior). Edge detection is quite useful in an application of this kind. It was highly useful in the case of the skull data, but finding edges that form a circle (confer Hough transform) as in a skull is somewhat of a simplified problem. Developments should aim to address many such issues.

[47] Recent discussions also suggest that data may be similar to that used in Davies' thesis.

[48] Warps placement truly seems tactless and poor at present, but this needs to be confirmed by evidence.

7. Challenging Issues

There are several issues that cannot be ignored and should therefore be systematically listed. Here is a brief unordered list of issues that appear to induce uncertainties and confusion:

(1) A sequential series of warps is often an expensive step that results in poor productivity.
(2) There is a wide range of warps and there is no consent on which the most effective ones are.
(3) The existing algorithms are very slow and require long periods of waiting time until constructive feedback is received.
(4) An existing system that sometimes struggles with one-dimensional data is required to be extended to 2-D and preferably 3-D too.
(5) The data handled by the existing approach tends to be excessively simple. Feasibility of such an approach in complex applications is still unknown.
(6) Medical imaging requires high fidelity and reliability. Unfortunately, the output from inner-body imaging has a significantly low SNR[49]; this conflicts with the fidelity requirement. The accuracy of this approach, e.g. the establishment of correspondence, is then unsatisfactory for some of the more critical procedures.
(7) There is often little knowledge about the structures in an image and random warps are then the only reasonable choice, resulting in a slow process. Solutions might come up in the form of bottom-up analysis of an unknown image.
(8) Unanimous choice of warps type and choice of default complexity for the warps is missing. Therefore, uncertainty looms over the real performance capabilities, group-wise optimisation being a main concern.

It is expected that many of the issues above will wind up being taken into consideration. They may affect the feasibility of the project, lead to failures or halt the pursuit for the original aims[50]. The next section proceeds to outlining the plan that will be adhered to throughout the forthcoming year.

[49] The signal-to-noise ratio in medical images can be lower by orders of magnitude in comparison with visual images.

[50] Section 7 was bound to take a pessimistic point-of-view to describe worst-case scenarios. A more optimistic contemplation would have discussed the obtainable goals and the factors that make these goals hard to reach.

8. WORK PLAN

8.1. **Aims.** The aims of the project were formally specified in **Form 2** (see Appendix B) although many of them are not yet definite. From a nonspecific perspective, it is expected that:

(1) A full reproduction of past experiments should be trivial and extensions prevised.
(2) Development of existing code will commence to ultimately build genuine software.
(3) Difficulties should be identified to avoid future impasse.
(4) New practicable experiments should be set and their results recorded.
(5) Comparative figures will show the advancements of new methods.
(6) Critical evaluation of existing work and proposition of new methods will hopefully emerge.

8.2. **Milestones.** In line with **Form 2**, but from a broader, more formal scope, here are some very rough estimates of expected milestones. The following chart summarises some of the milestones expected and the corresponding deadline.

	Recommended completion date
Existing code mastered	December 15th, 2003
Submission of **Literature Report**	December 19th, 2003
Extension to code completed	December 27th, 2003
Presentation	N/A
Resolving Project Plan	January 2004
Progress Report Submission	March 1st, 2004
First implementation working	April 2004
Implementation entirely documented	July 2004
Experiments performed	August 2004
Continuation Report Viva completed	August 25th, 2004

Figure 11: General time guidelines

	Deadline date
Existing code mastered	December 20th, 2003
Submission of **Literature Report**	December 22nd, 2003
Extension to code completed	January 5th, 2004
Presentation	January - February 2004
Resolving Project Plan	January 2004
Progress Report Submission	March 24th, 2004
First implementation working	April 2004
Implementation entirely documented	July 2004
Experiments performed	August 2004
Continuation Report Viva completed	September 1st, 2004

Figure 12: Deadlines

The Gantt chart below attempts to assure compliance with the deadlines and guarantee that progress will be made as anticipated.

	Dec/03 **May/04**	**Jan/04** **Jun/04**	**Feb/04** **Jul/04**	**Mar/04** **Aug/04**	**Apr/04** **Sep/04**
Literature Report	√				
Presentation	√	√	√		
Work on existing code	√	√	√		
New implementation	 √	 √	√ √	√	√
Documentation	 √	 √	 √	 √	√
Experiments	 √	 √	 √	 √	
Continuation Report Viva			 √	 √	 √

Figure 13: Progress Gantt Chart

Work division that is project-specific and a better summary of the *technical* aspects will be entirely left out. More accurate aims are formulated in Appendix B for completeness.

8.3. **Contingencies.** As some feasibility considerations are yet to be resolved, it is vital that alternative directions for this research are realised and suggested. One facet of this issue is concerned with times at which *evaluation* of progress, development and achievement ought to be approximated, apart from the formal evaluation in April 2004. By recognising dead-ends at the earlier stages of work, wasted effort can essentially be avoided. There are several types of problems that can come up:

(1) A field is yet too poorly understood and there is a lack of basic knowledge to rely upon.
(2) Effort is already invested in the exact same field or problems that the project poses are found to be resolved already.
(3) The code dealt with is too hard to cope with.
(4) Given algorithms or conventional methods are too slow to work with productively[51].
(5) Alternative solutions with greater potential are identified, thereupon requesting all attention to be diverted to them exclusively.
(6) Experiments fail to produce the results expected.
(7) Progress is held back by time restrictions.

Obstructions which are prone to happen more frequently would be 4, 5 and even quite frustratingly 6.

[51]Frequently it appears to be the case that in order to get reasonable results, high computational power is mandatory. In the absence of this power, experiments might fail or become impractical.

9. SUMMARY

9.1. **Brief Overview.** Some of the main relevant concepts and techniques in existence have been explained and numerous examples have been given, although their number was restricted to allow for a broader survey. Most such techniques directly relate to the problems which need to be tackled and their utilisation in past and present has been thoroughly explained. As future experimentation is expected to rely on recent research and is most likely to involve similar ideas, algorithms and paradigms, continuous reading of technical reports, alongside reproduction of the experiments, will be an essential portion of the research approached.

The problems with current techniques were found to vary from the interest in efficiency to possible flaws and gaps, a part of which being driven by insufficient correctness arguments and lacking ground-truth. Without a doubt, there are phases in current research where heuristics take over on the expense of valid implementation that can be reasoned straight-forwardly. Many areas are still controversial and common assent is missing and might never be reached. As instances for the aforesaid claims, a group-wise brain analysis algorithm devised a wide range of domain-specific facts. Moreover, a major undecided issue is the most advantageous warp type and its corresponding complexity that strives to give ideal results per permanent time unit.

This project, much like other projects in this area, attempts to find some answers to the questions raised and resolve uncertainties and disagreement by providing empirical proof or implementing a convenient tool for quickly evaluating and profiling different ideas and approaches. Whether it will be successful in the sense that it should provide inarguable answers and discover new techniques that are ingenuous, this project should draw conclusion regarding performance, feasibility and validate or invalidate some results of previous work. Preferably it should surpass previous work that it shall build upon. No results will be taken for granted and a *critical* approach will be dispensed at all times.

Within the first year of the project it is hoped that an implementation of a warp and model testing prototype will be available to achieve dense correspondence across a set of synthetic (and hopefully medical) images in 2-D and 3-D. Software should be capable of looking into the behaviour of warps regardless of the nature and scale of the data. It must also respond within sensible time period, although the notion of suitable time period is loosely-defined. There is a growing belief that such tool can be of great interest to these who use and facilitate active appearance models. Nonetheless, there is a real snag as the data under consideration should be modified to approve the successful application of the techniques to data of higher dimensionality. The run-time and the results that can be retrieved within a limited time-frame is then the main impediment.

Although a partial time-line was specified for this project and its intermediate objectives, it is not yet clear where the project will turn and what it will eventually accomplish successfully. It is known, however, what should *ideally* be accomplished. Semiannual reports and documents will clarify the emerging plans and intentions as they become more concrete.

9.2. **Conclusions.** The appearance models currently used are not ideal in any sense and a solution to this flaw would be highly desirable. While it is not clear how to optimise models or how to evaluate a model [44], there are measurable means for arguing about the quality of these models comparatively. Amongst the main problems that are ordinarily seen in appearance models is their inferior performance, although this depends on the functionality required. Automation could have a significant contribution to such a model, but correspondence needs to be achieved first. Luckily, issues of correspondence have been investigated largely in the past decade so this should not be a peril. What is worth investigating even further is the ability of warps to improve models and at the same time encapsulate several analysis steps together. What is even more reassuring is the proven ability of models to improve warps selection and improve on existing group-wise registration methods. This improvement relies on the fact that a large-scale collective analysis replaces the weaker yet computationally inexpensive pair-wise scope.

9.3. **Final Discussion.** As described in Section 8, the project is now expected to continue along the lines of implementation with some reliance on the work and results produced by Katherine Smith. If difficulties often recur, there are various measures that can be taken to ensure productive alternatives are chosen. The next goal is to obtain dense correspondence across 3-D biomedical data using automatic self-instructing algorithms. As the project is intended to explore a scarcely known field, caution will be taken when time is spent without much potential on the horizon.

10. Appendix A: Primary On-Line Resources

As much of the reading was based on educational and personal sites from the World Wide Web, several dominant sources must be acknowledged.

- `http://www.math.ufl.edu/help/matlab-tutorial/`

  An extensive MATLAB tutorial from the University of Florida.

- `http://www.doc.ic.ac.uk/~dr/`

  Daniel Rueckert's academic pages.

- `http://www-ipg.umds.ac.uk/d.hill/`

  Derek Hill's abstracts and publications.

- `http://www.dcs.gla.ac.uk/~mc/`

  Technical Reports of Matthew Cairn.

- `http://www.imm.dtu.dk/image/research/`

  Related research in the Technical University of Denmark.

- `http://www.ai.mit.edu/~viola/`

  Pages maintained by Paul Viola.

- `http://www.isbe.man.ac.uk/~bim/`

  Publications and resources from Tim Cootes.

- `http://www.cs.jhu.edu/~wolff/course600.461/`

  Computer Vision at Johns Hopkins University.

- `http://www.cs.wisc.edu/~dyer/cs766.html`

  Computer Vision at the University of Wisconsin.

11. Appendix B - Form 2

<u>Date of meeting:</u> December 15th, 2003.

RESEARCH AIMS & OBJECTIVES

Synopsis of Research Project and overall aims:

- Fully automate obtaining a set of dense correspondences across a set of 3D medical images as a basis for building statistical models of shape and appearance.
- Develop a new approach with a rigorous theoretical basis and compare its performance with existing approaches to the problem.
- Apply the method(s) developed to demonstrate changes in morphology due to disease (or other causes) in a large dataset (eg brain, knee etc).

Objectives of research project for first year (full-time Students) or first two years (part-time Students), including literature searching:

- Establish benchmark results for correspondences obtained using existing non-rigid registration algorithms.
- Develop an in-depth understanding of the literature on: non-rigid registration, active shape/appearance models, and minimum description length methods (generally and as applied to shape correspondence).
- Develop a general understanding of current methods and problems in computer vision, with particular emphasis on medical image analysis.
- Carry out initial experiments using synthetic data to gain an insight into the problem of automatic image correspondence and an understanding of the key problem areas.
- Obtain and analyse initial results using both synthetic and real data (possibly only 2-D).
- Develop a plan for future work, based on the experience of the first year.

Key objectives for first 3 months:

- Complete machine learning module successfully.
- Establish a pattern of background reading.
- Undertake a detailed review of the literature in non-rigid registration, active shape/appearance models, and minimum description length methods (generally and as applied to shape correspondence).
- Gain good familiarity with using MATLAB to run computer vision experiments and to analyse results.
- Establish simple 1-D model building framework using MATLAB software from Kate Smith.
- Plan presentation for student seminar.

Key objectives for first year:

- See objectives for first year above.

VARIOUS COURSES

Course/Seminar Title	**Dates (if known)**
Introductory Course	22/09/03 - 26/09/03
Library Visits	ISBE Research Library: perpetual
Regulatory Core Courses	1/10/03 - 1/06/03
Computing Skills and Statistics	N/A
1st Year Workshop	N/A
Health and Safety Training (Compulsory)	N/A

Acknowledgement. Many thanks to Mark O'Leary for allowing me to proceed with all my academic duties while at work and to Chris Taylor for longer-than-expected meetings and relentless support. The project is funded by the EPSRC.

REFERENCES

[1] S. C. Joshi and M. L. Miller, "Landmark Matching via Large Deformation Diffeomorphisms," IEEE Transaction on Image Processing, vol. 9, pp. 1357 – 1370, 2000.

[2] Y. Wang and L. H. Staib, "Integrated approaches to non-rigid registration in medical images," presented at Workshop on Applications of Computer Vision, 1998.

[3] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, J. Maurer, C. R., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawkes, C. Studholme, J. B. A. Maintz, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, J. Maguire, G. Q., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson and R. P. Woods, "Comparison and evaluation of retrospective intermodality brain image registration techniques," Journal of Computer Assisted Tomography, vol. 21, pp. 554–566, 1997.

[4] M. Xu and W. L. Nowinski, "Talairach-Tournoux Brain Atlas Registration Using a Metalforming Principle-Based Finite Element Method," Medical Image Analysis, vol. 5, pp. 271 – 279, 2001.

[5] L. Younes, "Deformations, Warping and Object Comparison: A Tutorial," 2000.

[6] M. Mitchell Tom, Machine Learning. New York ; London: McGraw-Hill, 1997.

[7] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton and C. J. Taylor, "3D statistical shape models using direct optimisation of description length," in Computer Vision - ECCV 2002 Pt Iii, vol. 2352, Lecture Notes in Computer Science, 2002, pp. 3-20.

[8] R. H. Davies, C. J. Twining, P. D. Allen, T. F. Cootes and C. J. Taylor, "Shape discrimination in the hippocampus using an MDL model," in Information Processing in Medical Imaging, Proceedings, vol. 2732, Lecture Notes in Computer Science, 2003, pp. 38-50.

[9] G. J. Edwards, C. J. T. Taylor and T. F. Cootes, "Interpreting face images using active appearance models," presented at Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998.

[10] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active appearance models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 681-685, 2001.

[11] M. B. Stegmann, B. K. Ersboll and R. Larsen, "FAME - A flexible appearance modeling environment," IEEE Transactions on Medical Imaging, vol. 22, pp. 1319-1331, 2003.

[12] T. F. Cootes, C. Beeston, G. J. Edwards and C. J. Taylor, "A unified framework for atlas matching using Active Appearance Models," in Information Processing in Medical Imaging, Proceedings, vol. 1613, Lecture Notes in Computer Science, 1999, pp. 322-333.

[13] I.T. Joliffe, “Principal Component Analysis,” Springer Series in Statistics, Springer, New York, 1986.

[14] C. J. Twining and C. J. Taylor, "The use of kernel principal component analysis to model data distributions," Pattern Recognition, vol. 36, pp. 217-227, 2003.

[15] J. Ahlberg and R. Forchheimer, "Face tracking for model-based coding and face animation," International Journal of Imaging Systems and Technology, vol. 13, pp. 8-22, 2003.

[16] D. W. Hansen, M. Nielsen, J. P. Hansen, A. S. Johansen and M. B. Stegmann, “Tracking eyes using shape and appearance,” IAPR Workshop on Machine Vision Applications, pp. 201-204, 2002
[17] M. Sonka, V. Hlavac and R. Boyle, Image processing, analysis and machine vision. Pacific Grove, Calif. ; London: PWS Publishing, 1999.
[18] T. F. Cootes, G. V. Wheeler, K. N. Walker and C. J. Taylor, "View-based active appearance models," Image and Vision Computing, vol. 20, pp. 657-664, 2002.
[19] T. F. Cootes, G. V. Wheeler, K. N. Walker and C. J. Taylor, “Coupled-view active appearance models,” British Machine Vision Conference, vol. 1, pp. 52-61, 2000.
[20] A. Lanitis, C. J. Taylor and T. F. Cootes, "Automatic Face Identification System Using Flexible Appearance Models," Image and Vision Computing, vol. 13, pp. 393-401, 1995.
[21] A. Lanitis, C. J. Taylor and T. F. Cootes, "Toward automatic simulation of aging effects on face images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 442-455, 2002.
[22] A. Lanitis, “PROSOPO - A face image synthesis system,” Advances in Informatics, Lecture Notes in Computer Science, vol. 2563, pp. 297-315, 2003.
[23] A. D. Brett and C. J. Taylor, "A method of automated landmark generation for automated 3D PDM construction," Image and Vision Computing, vol. 18, pp. 739-748, 2000.
[24] A. Hill, C. J. Taylor and A. D. Brett, "A framework for automatic landmark identification using a new method of nonrigid correspondence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 241-251, 2000.
[25] K. N. Walker, T. F. Cootes and C. J. Taylor, "Determining correspondences for statistical models of appearance," in Computer Vision - ECCV 2000, Pt I, Proceedings, vol. 1842, Lecture Notes in Computer Science, 2000, pp. 829-843.
[26] J. V. Hajnal, D. L. G. Hill and D. J. Hawkes, Medical image registration. Boca Raton, Fla. ; London: CRC Press, 2001.
[27] D. Rueckert, A. F. Frangi and J. A. Schnabel, "Automatic construction of 3D statistical deformation models using non-rigid registration," presented at MICCAI, 2001.
[28] C. J. Twining and S. Marsland, "Constructing diffeomorphic representations of non-rigid registrations of medical images," presented at IPMI, 2003.
[29] J. Lötjönen and T. Mäkelä, "Elastic Matching Using a Deformation Sphere," presented at Fourth International Conference on Medical Image Computing and Computer Assisted Intervention, 2001.
[30] J. F. Kenney and E. S. Keeping, “Mathematics of Statistics,” part 2, 2nd edition, Princeton, New Jersey: Van Nostrand, 1951.
[31] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," International Journal of Computer Vision, vol. 24, pp. 137-154, 1997.
[32] J. P. W. Pluim, J. B. A. Maintz and M. A. Viergever, "Image Registration by Maximisation of Combined Mutual Information and Gradient Information," IEEE Transaction on Medical Imaging, vol. 19, pp. 809 – 814, 2000.
[33] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information," IEEE Transactions on Medical Imaging, vol. 16, pp. 187 – 198, 1997.
[34] B. Likar and F. Pernus, "A Hierarchical Approach to Elastic Registration Based on Mutual Information," Image and Vision Computing, vol. 19, pp. 33 – 44, 2001.

[35] C. Studholme, D.L.G. Hill and D.J. Hawkes, “An overlap invariant entropy measure of 3D medical image alignment,” in Pattern Recognition, vol. 32, issue 1, pp. 71-86, 1999.

[36] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, “Multimodality image registration by maximization of mutual information,” in IEEE Transactions on Medical Imaging, vol. 16, issue 2, pp. 187-198, 1997.

[37] J. P. W. Pluim, J. B. A. Maintz and M. A. Viergever, “Mutual information matching in multiresolution contexts,” in Image and Vision Computing, vol. 19, issues 1-2, pp. 45-52, January 2001.

[38] W. H. Press, Numerical recipes in C/C++ : the art of scientific computing. Cambridge: Cambridge University Press, 2002.

[39] S. K. Warfield, J. Rexilius, P. S. Huppi, T. E. Inder, E. G. Miller, W. M. Wells, III, G. P. Zientara, F. A. Jolesz and R. Kikinis, "An Entropy Measure to Assess Nonrigid Registration Algorithms for Statistical Atlas Construction," presented at MICCAI, 2001.

[40] D. Rueckert, A. F. Frangi and J. A. Schnabel, “Automatic construction of 3-D statistical deformation,” IEEE Transactions on Medical Imaging, vol. 22, issue 8, pp. 1014-1025, 2003.

[41] J. R. Rissanen, "Stochastic complexity in statistical inquiry," presented at World Scientific Series in Computer Science, Singapore, 1989.

[42] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton and C. J. Taylor, "A minimum description length approach to statistical shape modeling," IEEE Transactions on Medical Imaging, vol. 21, pp. 525-537, 2002.

[43] S. Marsland, C. J. Twining and C. J. Taylor, "Groupwise non-rigid registration using polyharmonic clamped-plate splines," presented at MICCAI, Montreal, Canada, 2003.

[44] A. C. W. Kotcheff and C. J. Taylor, "Automatic construction of eigenshape models by genetic algorithm," in Information Processing in Medical Imaging, vol. 1230, pp. 1-14, Lecture Notes in Computer Science, 1997.

[45] A. Najmi, R. A. Olshen and R. M. Gray, "A criterion for model selection using minimum description length," in Proceedings Compression and Complexity of Sequences, pp. 204-214, Salerno, Italy, 1997.

[46] C. J. Beeston and C. J. Taylor, "Automatic landmarking of cortical sulci," in Medical Image Computing and Computer-Assisted Intervention - MICCAI 2000, vol. 1935, Lecture Notes in Computer Science, 2000, pp. 125-133.