Second Year Progress Report

ROY S. SCHESTOWITZ
IMAGING SCIENCE AND BIOMEDICAL ENGINEERING
UNIVERSITY OF MANCHESTER
UNITED KINGDOM

1. Introduction

THIS report explores progress made by the author in the past 2 years with particular emphasis on the second year. The earlier sections briefly cover some background details, while the latter outline points of progress with contextual reference to the background. Lastly, §A takes an overview and lists possible ways of assembling these points of progress as to make them more cohesive.

2. Image Registration

Image registration, or more particularly *non-rigid* image registration (NRR), is an essential pre-processing step that enables easier interpretation of image sets. Registration of images is a task which involves repeated image transformations, ultimately aiming to make a group of images look more similar. By striving to and thus achieving cross-image similarity, change can be more easily analysed and image structures better understood. Various approaches exist to solving the NRR problem. These different approaches vary in terms of their representation of transformations, the notion of similarity among images, and the method by which good transformations get selected.

In most cases, registration is posed as a pair-wise problem, whereby several images are transformed to fit one single image, commonly known as the reference or template image. A registration that takes into account the entire set and treats images equally, on the other hand, could and should result in a better overall registration.

Due to the arbitrary choice of a reference image, a pair-wise approach is rarely constrained and its results are greatly affected by that subjective choice of a reference. Put differently, depending on which image gets selected as the reference, different results are to be reached. Moreover, the loosely-defined approach by which registration gets solved means that it is difficult to reason about correctness of the result or even quantify it reliably.

In order to assess the power of registration algorithms 'off-line', one can make use of the ground-truth solution, often straying away from that solution by applying perturbation. It is then possible to investigate an algorithm's ability to 'annul' the effect of the perturbation by performing registration. A good registration algorithm will be able to recover the ground-truth solution. The drawbacks of such an approach are that ground-truth solutions must be provided *a priori* and partial recovery of the solution needs to be evaluated somehow.

3. Synthesis and Registration

Appearance model, which are further discussed in §4, can be used for interpretation by synthesis. Given a set of images, models can be built that capture the observed variability in the set. Any set of images, from which an appearance model is built, has an important trait: the better-registered the set, the better its model. In other words, when images are aligned properly and look similar, the model itself improves.

Since models as such are generative, they can also be used 'in reverse', synthesising images that they describe. As explained in §6, there is an empirically-justified method for evaluating models, transitively inferring the quality of registration. As a matter of fact, it is our contention that modelling and registration are inherently an identical problem. The remainder of this report explains how models can be evaluated, thereby helping us evaluate non-rigid registration as well.

Looking at our recent work more comprehensively, registration can properly be evaluated by creating models and evaluating these. Furthermore, evaluation of such models can assist registration or even serve as the sole criterion that drives registration. More on this individual aspect of our work is described briefly in §10.

4. Models of Shape and Appearance

Statistical models of shape and appearance, which are often referred to as *combined* appearance models, encapsulate variation across the set from which they are built. To construct these models, one needs to establish dense point-to-point correspondence across all images in the set, thereby highlighting analogous structures.

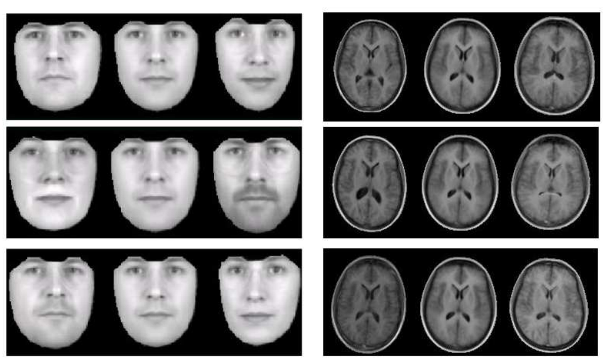

Fig. 1. Appearance models showing the effect of varying the first, second, and third model parameters. Each of the models is subjected to variation of at most ± 2.5 standard deviations.

Construction of appearance models involves a stage where variation in shape and texture (image intensities) are learned in turn. Shape can be represented as a vector $\mathbf{x}$ while texture represented as a vector $\mathbf{g}$. Both shape and texture can be directly controlled by models of the form

$$\begin{array}{l}\mathbf{x} = \overline{\mathbf{x}} + \mathbf{P}_s \mathbf{b}_s \\ \mathbf{g} = \overline{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g\end{array} . \tag{1}$$

In the formulation above, $\mathbf{b}_s$ are the shape parameters, $\mathbf{b}_g$ are the texture parameters, $\overline{\mathbf{x}}$ and $\overline{\mathbf{g}}$ are the mean shape and texture while $\mathbf{P}_s$ and $\mathbf{P}_g$ are the principal modes of shape and texture variation respectively. In practice, there is a tight correlation between shape and in intensity so a combined statistical model of the form

$$\begin{array}{l}\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \\ \mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c}\end{array} \tag{2}$$

appears to work even more gracefully, integrating both sources of variation. The model parameters $\mathbf{c}$ control both shape and texture simultaneously and $\mathbf{Q}_s$, $\mathbf{Q}_g$ are matrices describing the modes of variation derived from the training set.

5. Tolerant Similarity Measure

Further to an explanation in §2, which makes a mentioning of image similarity, we seek a measure of similarity that is robust to slight localised differences between images. If an image gets treated merely as an ordered collection of pixels, then strict pixel-to-pixel difference can be derived, which is Euclidean. This does not, however, account for any spatial relationship between pixels. A slight mismatch can entail a significant penalty, which is uncalled for. Consequently, the sensitivity of our similarity measure to change is so high that the Euclidean measure becomes almost meaningless.

A better method of computing image similarity should take into account additional spatial properties (e.g. movements, bends and elasticity). This seems rather natural once we realise that corresponding pixels across the images may lie in the vicinity of geometrically corresponding points in the images. As a result we are then inclined to use the shuffle distance, which can cover a large disc-shaped region where the best match for any given point is to be identified (see Fig. 2). The effects of varying the 'scope' of the shuffle distance are shown in Fig. 3.

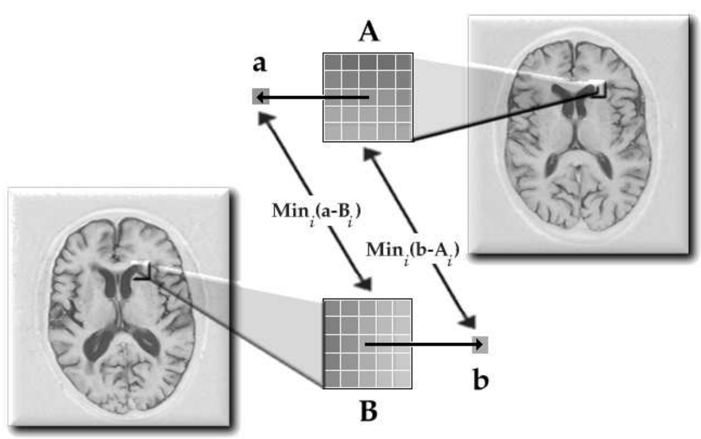

Fig. 2. The symmetric calculation of a shuffle difference image. Each pixel is compared to its closest match in the other image.

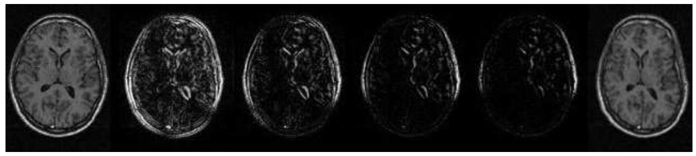

Fig. 3. Surveying image distances for evaluation of brain registration. On the left and the right: distinct yet similar images; Across the centre (from left): shuffle distance with $r = 0$ (Euclidean distance), 1.5, 2.9 and 3.7.

6. Specificity and Generalisation

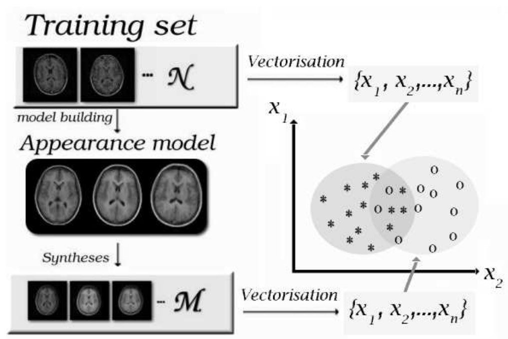

Fig. 4. The framework of model evaluation where a model is constructed from the training and images are generated from the model. Each image is vectorised and can then be visualised as a cloud in hyperspace. The rationale behind Generalisation and Specificity is made clearer in the context of clouds where overlap of the clouds can be measured trivially.

A valuable approach to the assessment of combined model involves generation of images from that model and then a computation of how well they fit the model and vice versa. Images are created by taking the mean image from the model and selecting random values for the parameters **c**, which in turn enables us to generate a large number of synthetic reconstructions from the model.
As correspondences from which the model gets built degrade (equivalent to mis-registration), the resulting model becomes less capable of reconstructing images of the same type. The model is less able to generate valid images that are not in the training set and struggles to only synthesise new images similar to those in the training set. The former property we refer to as Generalisation and the latter – Specificity. They 'tighten' the model to the data that it represents. If we decide to embed training images and model-synthesised images in a very high dimensional space, then we wish to obtain a model, represented by a synthetic images cloud, that best overlaps a cloud which is formed by the model's training set (see Fig. 4).

7. Perturbation Framework

In order to test our method, it is required that we take ground-truth registrations and transform them as to degrade them in a well-understood fashion. By doing so, a set comprising different registrations of varying quality can be obtained. By applying our method to each such set in turn, we should be able to demonstrate that our method discerns good registrations from worse ones. Ideally, our method would be able to provide a benchmark whose scale is monotonic and span a wide range of possible mis-registrations.

To perturb the registration we use clamped-plate splines[1] which are composed of 25 knot-points, all of which are randomly distributed across the images. By increasing the magnitude of the warp, we can directly increase the level of mis-registration.

8. Method Validation

Using the perturbation method outlined succinctly in§7, we can measure the quality of different models. These models were obtained from the same the registrations with varying extents of perturbation applied,

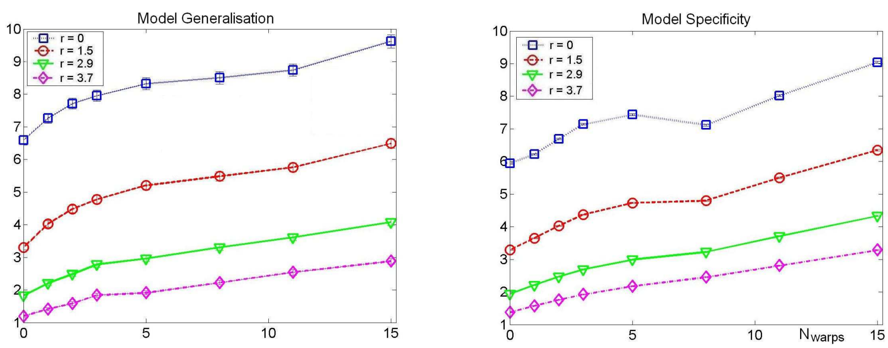

Fig. 5. An older evidence that Generalisation and Specificity of brain models degrades as their registration degrades

The results of an experiment which tests the effect of increasing mis-registration are shown in Fig. 5. The curves indicate that, whichever shuffle 'scope' we choose, both Specificity and Generalisation increase, which means that they get worse as registration gets artificially degraded. These results which include a variety of shuffle distances show a consistent trend, which is an encouraging behaviour. It implies that we needn't compare each pixel against a very large region, hence efficiency if improved. Furthermore it shows that there is little dependency on yet another parameter, which is the region size for the shuffle distance.

9. Assessment of Non-Rigid Registration using Models

Having argued for the validity of the method, we proceeded to a comparing of different methods of NRR. Of particular interest were group-wise and pair-wise method. Much as we had hope to show, it turned out that group-wise registration is superior to pair-wise. This assumption even holds over a wide range of parameter values that can be selected. This led to the much anticipated conclusion that our method can evaluate registration algorithms and little or no tweaking should be involved in advance so the method may be generic.

[1] The clamped-plate splines (CPS) are based on Green's function and are defined by their radius, magnitude and direction.

Current work appears to indicate that results of our evaluation agree with results that are based on ground truth. The advantage of our approach, on the other hand, is that is is data-driven and does not require additional information to be provided.

10. Model-based Registration

At an earlier stage of this project, yet another idea which was based on the relationship between registration and models, got investigated in practice. Sparked by successful work on shape models, wherein good models were constructed by minimisation of their complexity, a similar approach was applied to combined models, i.e. models incorporating both shape and intensity. Images were transformed in a well-controlled manner, their model was built and transformations were subsequently accepted or rejected based on the quality of the model.

Despite the impracticality of the algorithm, which suffered from heavy computational requirement, proof-of-concept results were quite satisfactory.

11. Summary and Conclusions

We have devised and shown the practicality of a model-based approach to evaluation of non-rigid image registration. To carry out such an evaluation, no ground truth needs to be involved and yet accuracy of the method is comparable with that of methods that exploit known solutions. To show that the method is both practical and accurate, we performed a validation by perturbing the ground-truth and confirming that our evaluation method detects the degree of perturbation.
We then proceeded to evaluating various registration algorithms, most notably the pair-wise and group-wise approach and showed that better results are obtains when the latter gets used.

Thus approach which was developed in the second year is an important advancement, which makes NRR assessment an objective process that does not require ground-truth solutions. The method can therefore be applied readily to evaluate and compare a variety of registration methods, just as it was able to assess and distinguish between pair- and group-wise registration.

In addition to the method which is able to assess registration, we have also developed a model-base objective function, which takes our evaluation criterion and uses it to guide registration. While the approach seems to work in theory, it is rather computationally-expensive and is therefore impractical to consider in 2- and 3-D.

A. Speculated Thesis Structure

Below lies a *narrow*[2] thesis 'skeleton', which may or may not form a valid starting point for the composition of the thesis body. Moreover, it aims to proliferate ideas and encourage discussion.

- ➤ Introduction
- ➤ Models
  - ✦ Statistical Models
  - ✦ PCA
  - ✦ Model Construction
  - ✦ Shape Models
  - ✦ Intensity Models
  - ✦ Combined Models
- ➤ Non-rigid Registration
  - ✦ Warps
  - ✦ Similarity
  - ✦ Group-wise versus Pair-wise
- ➤ MDL Shape Models
- ➤ Model-based Registration
  - ✦ MDL objective function
  - ✦ Results
- ➤ 3-D Registration Framework
- ➤ Automatically Building Appearance Models
- ➤ Evaluating Models and Non-rigid Registration
  - ✦ Specificity and Generalisation
  - ✦ Results
  - ✦ Perturbation Framework
  - ✦ Comparison with Overlap-based Evaluation
- ➤ Future Exploration
- ➤ Summary, Discussion and Conclusions

[2] The bulletpoints must not be treated as possible sections and subsections. There are merely streams of consciousness that can later be sub-divided and/or expanded.

B. Events and Activities

B.1. First Year

Date	Description
22-26 September	Introductory week
6-24 October 2003	Machine learning
20 January 2004	Oxford plenary meeting
10, 17 March 2004	Thesis writing seminar
2, 25 March 2004	Student presentations
30-31 March 2004	Manchester plenary meeting
April 2004	Mathematical methods
27 May 2004	UCL S&F meeting
2 June 2004	Ph.D. Workshop
21-25 June 2004	Surrey Summer School
13 July 2004	S&F Workshop, Gordon Museum

B.2. Second Year

Date	Description
19-22 September 2004	MIUA Summer School, Imperial College
19-20 October 2004	UCL plenary meeting
January 2005	Student presentations
26-27 January 2005	Manchester plenary meeting
22 April 2005	UCL plenary meeting
19-20 May 2005	2nd Year Ph.D. Workshop
24-25 May 2005	Oxford plenary meeting
21 July 2005	Oxford Structure and Function plenary meeting

Also see Form 11 where compulsory activities are listed.