Groupwise objective function for automatic model-building

Kate Smith

November 21, 2003

1 Aims and motivation

We think we can improve on existing pairwise non-rigid registration algorithms by incorporating information derived from the set of images being registered, rather than by choosing an arbitrary image as reference, or a very simple statistically generated reference such as the mean image. Instead we wish to use a richer statistical description (a model) of the group as our reference. The model we will use is an appearance model, which is formed from a combination of a shape model and an intensity model (see Tim's notes). We will evaluate each warp of the image being registered according to whether the model built using the warped image is more or less compact than a model built from the unwarped image.

The reasons why we think this might improve matters are that given ambiguous correspondences (and in any non-rigid registration problem there are infinite degrees of freedom, and infinite possible answers) and no principled method of choosing the correct answer it will converge to the most probable. In practical terms, we hope that this method will produce less complex, noisy warps than other methods - the simplest warp that explains the data.

Following from previous work in the department on information theoretic measures, we considered an objective function based on the determinant of the covariance matrix (i.e. the product of the variances of the modes of the combined appearance model) of an appearance model built from the current estimate of the registered images. This value has been shown to be a good approximation to the minimum description length of the data (see Rhodri's thesis and Kotcheff and Taylor 1998).

However, there are clearly several issues with this objective function:

- The appearance model takes the form

$$\mathbf{b} = \begin{pmatrix} \mathbf{W_s b_s} \\ \mathbf{b_g} \end{pmatrix} \tag{1}$$

  , where $\mathbf{b_s}$ is the set of parameters of a deformable shape model and $\mathbf{b_g}$ is the set of parameters of a deformable intensity model (see Tim's notes)

Additionally, there is a parameter, $\mathbf{W_s}$, to adjust for the difference in units between the shape and intensity components of the vector. There is no principled way to determine the value of this parameter, as there is no correspondence understood between the units of the shape and intensity components, and this will affect the value of the objective function.

- This objective function may result in slow convergence of the algorithm - with a measure derived from the entire set of images, altering one image will only have a small effect on the value of the objective function, and smaller the greater the number of images.

- There will inevitably be a high computational cost - the model will have to be rebuilt frequently.

2 Data

To do a proof of principle for the key ideas we used extremely simple data - a 1D bump symmetrical about its centre and with varying width and height (i.e. two uncorrelated modes of variation). See figure 1 for some examples. The aim of our registration technique is to find the correspondences between the bumps by correctly aligning the step edges. We want minimal warping elsewhere, and we do not want to attempt to correct for the differing heights of the bumps.

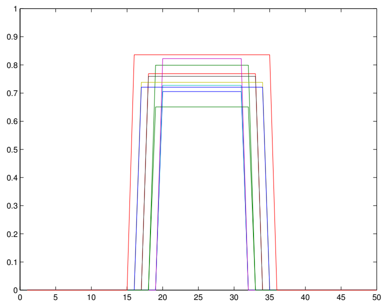

Figure 1: Plots of 10 bumps generated for test data.

3 Results

3.1 What does a model look like?

Figure 2 shows a model of the intensities of the initial misaligned images. The ragged edges are the result of the PCA decomposition of a misaligned step edge. The shape model of these images is trivial, as there has been no warping and all images have the same initial distribution of points. Figure 3 shows the combined model built from these images. It has only one mode which encompasses varying width and height of the bump. The central image shows the mean of the data, and the images to each side show the mean ± one or two standard deviations. The value of the determinant of the covariance matrix is 6.0607e-197.

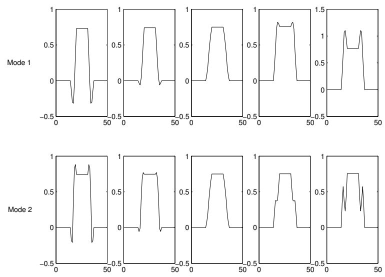

Figure 2: The first two principal components of an intensity model built from the unwarped images.

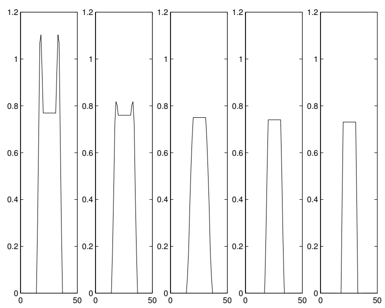

Figure 3: The single mode of an appearance model built from the unwarped images.

Figure 4 shows a model built from the images that have been perfectly aligned with a piecewise linear warp based on the known misregistrations. It has two modes of variation which represent orthogonal combinations of the varying height of the bump and the varying warps applied to the bump. The value of the determinant of the covariance matrix is 6.4566e-198, i.e. about one order of magnitude smaller than in the unregistered case.

3.2 Models built from misregistered data

To look at the behaviour of a model as the data approaches registration, I added varying amounts of error (in the form of randomly applied clamped plate spline warps) to the registered data. A misregistered model shows variation like that of the unregistered data above, with ringing in the intensity modes decreasing as the model approaches registration. Figure 5 shows the modes of variation of a model built from misregistered data.

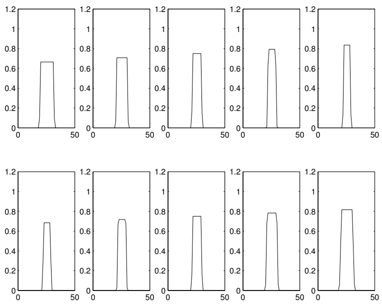

Figure 4: The two modes of an appearance model built from the perfectly aligned images.

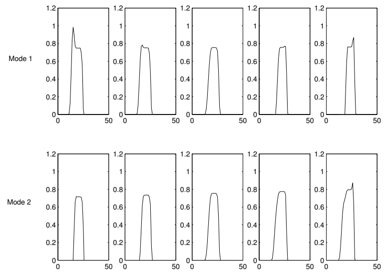

Figure 5: Two components of appearance model built from misregistered data.

Figure 3.2 shows the increases as misregistration increases in both the number of modes of variation of the shape and intensity models, and in the value of the determinant of the covariance matrix. We can conclude from this that the value of the objective function has a minimum at registration and and increases with greater degrees of misregistration and therefore may be something which is useful to optimise.

3.3 Choosing the best value of the shape/intensity weighting parameter

We need to choose a value for the parameter weighting shape and intensity ($\mathbf{W_s}$). Figure 3.3 shows how the score of the objective function built from the determinant of the covariance matrix with error, using different values for $\mathbf{W_s}$ - a selection of different constants and two data-derived values. The ratio of the total variances of the current shape and intensity models has proved an effective normalisation factor in the building of appearance models in the past, however we had doubts about whether it would be suitable in this case because it is dependent on the current estimate of the warps, therefore it might distort the measure when the images are still misregistered. We also tried using the mean magnitude of the edge strengths in the image, because this is independent of

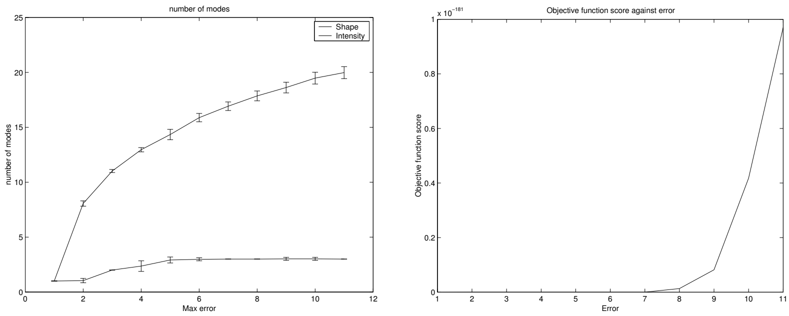

Figure 6: Increases in number of modes and objective function score as misregistration increases.

the current (probably incorrect) estimate of the warps.

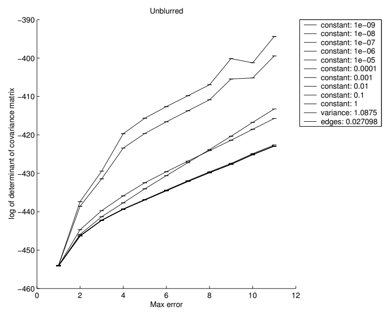

Figure 7: Increase in objective function score as misregistration increases, for different values of $\mathbf{W_s}$.

It can be seen from this that the objective function converges with all of the values tested. However, there does not appear to be any significant difference in the rate of convergence.

3.4 Specificity and generalisability

Because there are so many possible solutions to the correspondence problem and ground truth is subjective, we need to find ways to evaluate the resulting model independent of the objective function. For this purpose, when looking at shape models, Chris and Rhodri developed measures called 'specificity' and 'generalisability'. Specificity is defined as the distance from an example generated from the current model to the nearest example in the training set. Generalisability is defined as the distance from an unseen example to the approximation generated by optimising the parameters of the model. (For more detail, see Rhodri's thesis.)

As can be seen from figure 3.4, specificity increases with greater misregistration, as would be expected because a less compact model generated from more widely varying data will include a wider range of examples, including invalid examples which will be far from the valid examples in the training set.

However, no clear pattern can be seen in the graph for generalisability versus error (figure 3.4). This makes sense because a poor model is equally able to approximate valid examples as a good model - it is a poor model because it is also able to generate invalid examples.

Something it might be useful to measure is the time taken to find generalisability for more and less compact models - one would expect that a more compact model would converge much more quickly as it is searching only over the space of valid examples.

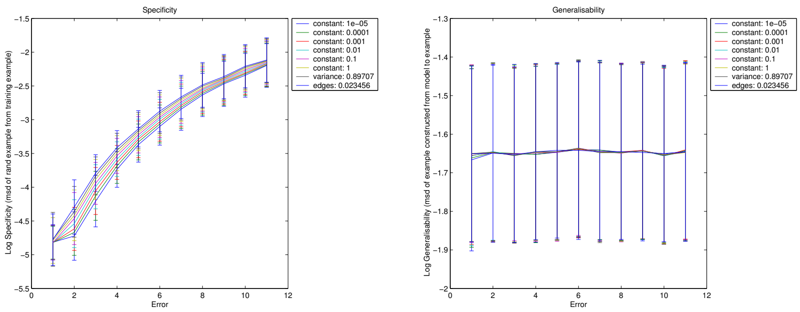

Figure 8: Specificity and generalisabiility scores as misregistration increases, for different values of $\mathbf{W_s}$.

Altering the value of $\mathbf{W_s}$ appears to make little difference, and no useful difference, to the behaviour of the specificity and generalisability measures. Figure 3.4 depicts the relationship between specificity and generalisability for different errors (the different coloured lines) and values of $\mathbf{W_s}$ (the different coloured dots on the lines), in the hope that we could see a value of $\mathbf{W_s}$ that produced smaller values for both specificity and genralisability, but it is hard to see any pattern in this data.

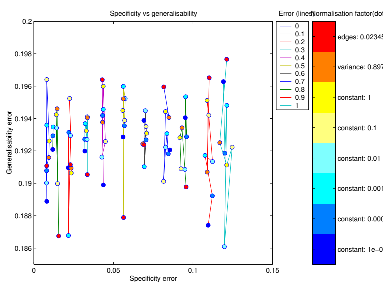

Figure 9: Relationship between specificity and generalisability measures for different errors and values of $\mathbf{W_s}$

3.5 Attempting to align data automatically

Given that the objective function behaves correctly as registration is approached, will it be a good driving force for an optimisation process? I have built models on the 1D bump data using two different objective functions:

- mean squared differences in data intensity, relative to a randomly chosen reference image (MSD)
- the model-based objective function using the determinant of the covariance matrix of an appearance model built from all the data to describe model quality

I have also used two different spline methods of warping, with several methods for controlling the placement of the splines/spline knotpoints:

- composition of several single point clamped plate splines
  - splines placed randomly
  - splines placed on a fixed grid
  - splines placed on strong image edges
- one multipoint clamped plate spline
  - knotpoints placed randomly
  - knotpoints placed on a fixed grid
  - knotpoints placed on strong image edges

Optimisation was also varied by both optimising all the warps being applied at the same time and optimising one at a time.

3.5.1 Automatic model-building algorithm

generate test images
for (set)

- for (iteration)
  - for (images)
    - add warp(s)/knotpoint(s) and warp image
    - evaluate objective function - either
      - (MSD) calculate MSD between this image and the reference; or
      - (Model-based) build a model from the current set of images, including the warped version of the image currently being optimised, and calculate the determinant of the covariance matrix of the model

Visually excellent results can be obtained with only one iteration of the optimisation using the MSD objective function (for example, figure **??**), however we require many more iterations to achieve results approaching the same quality when using the model-based objective function. I have used 50 (figure 3.5.1) and 25 iterations, between which two points the results clearly improve, but I have not experimented to find out how to determine the optimum number of iterations.

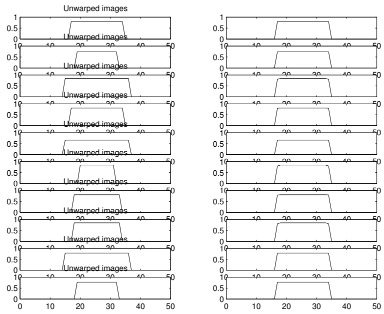

Figure 10: Original images (left) and warped images (right) after one iteration of registration using a multipoint CPS with knotpoints on edges with an MSD objective function

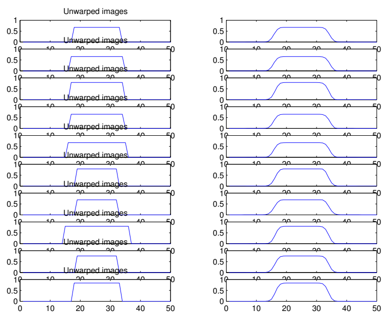

Figure 11: Original images (left) and warped images (right) after 50 iterations of registration using a multipoint CPS with knotpoints on edges with a model-based objective function

The most obvious difference that can be seen between the two objective functions is the time it takes to calculate them, much as was expected. The quickest of the model-based regimes takes of the order of 100x more time to run than does the equivalent MSD regime; the slowest requires days to run. The major factor influencing time taken to optimise, apart from the objective function, is the placement method - the more knotpoints/warps are used to achieve the final warp, the longer the optimisation takes.

There is a problem with using an objective function of which an grey-level difference-based measure forms a part (which is true of both objective functions tested here) in that the perfect warp does not necessarily give the best objective function score. A better score can be achieved by blurring out the edges of the bump, thus altering the value of the peak and giving better intensity matching at the cost of the correspondence quality. This will result in a much more complex warp and with some ideal objective function the benefit of the improved intensity matching might be balanced out by the cost of the warping to give the right result, but it is hard to see how this could be achieved in practice.

This type of problem is often handled by using a histogram-based measure such as mutual information. However using the multipoint clamped plate spline rather than a composition of single point splines also tackles this, as you can see from the relative sharpness of the edges in figures This can be seen in action if the single point CPS warps are allowed to run for many iterations - a better MSD score is achieved, however the model size becomes greater (need the data to show this, so don't take it for granted!).

From the figures below, it is hard to see any advantage conferred by the model-based objective function. It is orders of magnitude slower and both the intensity matches (as measured by the Mean MSD) and the models built from the registered data are of poorer quality. However an objective function such as MSD doesn't really have any future: it will certainly fail with multimodal or noisy data. This is where it is really becoming necessary to use some more complex data to test the objective functions. This test is also a very difficult one - it is all about approximating step edges, which is hard. The model-based objective function may perform better with more realistic data, without many step edges.

Warp placement	Mean MSD	Mean determinant of covariance	Mean shape modes	Mean time (s)
MSD objective function, joint optimisation, multi point CPS				
Edges	0.0024	1.5694e-72	2.72	2.6120
Grid	0.0035	5.1325e-61	6.5400	17.7324
Random	0.0043	9.0841e-64	6.8800	7.9224
MSD objective function, sequential optimisation, multi point CPS				
Edges	0.0025	4.3846e-71	2.6800	2.7148
Grid	0.0045	1.3130e-60	6.5000	17.3630
Random	0.0043	3.8153e-61	7.1000	8.1018
MSD objective function, joint optimisation, single point CPS				
Edges	0.0263	1.2182e-61	2.5800	1.1453
Grid	0.0037	8.9865e-63	6.1000	16.4211
Random	0.0078	3.0265e-58	6.8000	4.1029
MSD objective function, sequential optimisation, single point CPS				
Edges	0.0279	3.6520e-61	2.6800	0.8732
Grid	0.0601	1.5250-54	6.5600	19.0840
Random	0.0600	2.2001e-54	6.5200	20.7202
Model-based objective function, joint optimisation, multi point CPS, 25 iters				
Edges	0.0109	2.6444e-67	4.1200	213.7959
Model-based objective function, sequential optimisation, multi point CPS, 25 iters				
Edges	0.0083	5.0116e-69	3.9000	216.9614
Model-based objective function, joint optimisation, single point CPS, 25 iters				
Edges	0.0219	1.5159e-63	5.7200	92.7264
Model-based objective function, sequential optimisation, single point CPS, 25 iters				
Edges	0.0181	4.9082e-64	5.3600	105.8182
Model-based objective function, joint optimisation, multi point CPS, 50 iters				
Edges	0.0107	4.8950e-67	3.8400	393.3517
Model-based objective function, sequential optimisation, multi point CPS, 50 iters				
Edges	0.0081	2.4034e-69	4.0000	406.2829
Model-based objective function, joint optimisation, single point CPS, 50 iters				
Edges	0.0216	4.2290e-64	5.7600	169.5216
Model-based objective function, sequential optimisation, single point CPS, 50 iters				
Edges				

4 Conclusion and ways forward

I have shown that an objective function based on the determinant of the covariance matrix does have a minimum at registration and, in the case of this very simple data, does not have false minima on the way to registration. The optimum value of the parameter weighting shape and intensity components has not been identified, but it appears to make very little difference to the results of the optimisation.

As expected, the model-based objective function proved far more timeconsuming to calculate than a non-model-based measure. There were in fact no clear advantages to this method with this test set, but the test set was very unrealistic. The extremely simple nature of this dataset means that it is not obvious that the results here will extrapolate to real data. The objective function should be tested on more complex synthetic data and on real images. It should also probably be compared with other commonly used objective functions such as mutual information.

Using as few parameters as possible for the model makes a very big difference to the scale of the optimisation, so it would seem advisable to choose a low-dimensional warp representation, with the parameters chosen in some data-driven way.

5 Code

All experiments were carried out in Matlab, using a combination of code which can be checked out of the ISBE CVS repository (ask Mike Rogers or Gareth Jones for access) and code written by me which I will pass on to Roy. The functions to run to reproduce the results above are eval_1d_app_model_obj_fn, for the evaluation of the objective function with perfect registration and mis-registration, and build_1d_model, for running automatic registration.