Evaluation of Performance of Part-based Models for Groupwise Registration

Steve A. Adeshina
steve.adeshina@postgrad.manchester.ac.uk
Timothy F. Cootes
t.cootes@manchester.ac.uk

Imaging Science and Biomedical Engineering,
School of Cancer and Enabling Sciences,
The University of Manchester,
Manchester, UK.

Abstract

We evaluate the performance of a system which addresses the problem of building detailed models of shape and appearance of complex structures, given only a training set of representative images and some minimal manual intervention. We focus on objects with repeating structures (such as bones in the hands), which can cause normal deformable registration techniques to fall into local minima and fail. Using a sparse annotation of a single image we can construct a parts+geometry model capable of locating a small set of features on every training image. Iterative refinement leads to a model which can locate structures accurately and reliably. The resulting sparse annotations are sufficient to initialise a dense groupwise registration algorithm, which gives a detailed correspondence between all images in the set. We demonstrate the method on a much larger set of radiographs of the hand while comparing results with that of the earlier work, we achieved a sub-millimeter accuracy in a prominent group.

1 Introduction

Many forms of model can be constructed if we have accurate correspondences defined across a set of training images. However, obtaining such correspondences can be difficult and time consuming. In most early work on statistical shape models, for instance [2], the correspondences were created manually. More recently there has been considerable research into automated methods of achieving correspondence, such as from boundaries in 2D or surfaces in 3D (eg [4]), or more generally by directly registering images using non-rigid registration methods or 'groupwise' techniques [3].

In our earlier paper we tackled the problem of registering images of objects with considerable shape variation and multiple similar sub-parts. The key problem with such data is one of initialisation. A common approach to groupwise registration is to first find an affine transformation which gives an approximate solution, then perform non-rigid registration to an evolving mean to obtain more exact results [3]. Unfortunately, with the degree of variability exhibited in the hands, the affine stage is insufficient.

We use a parts+geometry model [5]. The local geometry can be used to efficiently select between multiple candidates for the parts. Donner *et al.* demonstrated how a sophisticated

parts + geometry model can accurately locate points in such images and how such a model can be constructed automatically from a set of images in which only one is manually annotated [5]. However, the method was only evaluated on a small set of 12 hand radiographs.

In this paper we show how a simple parts + geometry model can be learned from a large set of images using only one manually annotated image and how this can be used to initialise a groupwise registration algorithm, leading to dense correspondences [1]. We extend our earlier work to deal with 536 images (as opposed to 94). The key problem is the huge variation that exist in registering radiographs of children and young adults for automatic determination of skeletal maturity. This makes the original method perform less effectively.

In the following we describe the technique in tackling the inherent variation, demonstrate its use and evaluate it by comparing the results with the initial work [1].

2 Methods

2.1 Multi-Resolution Patch Models

Given one or more training images in which a particular region has been annotated, we can construct a statistical model of the region. We assume that the region is of fixed shape, but may vary in size and orientation. In the simplest case the region is an oriented rectangle or ellipse, centred on a point, $\mathbf{p}$ with scale s and orientation θ.

If $\mathbf{g}(\mathbf{t})$ are the intensities sampled from n pixels in the region with pose parameters $\mathbf{t} = \{\mathbf{p}, s, \theta\}$, normalised to have a mean of zero and unit variance, then the quality of fit to a model is evaluated as

$$f_i(\mathbf{g}(\mathbf{t})) = \sum_{j=1}^{n} |g_j - \bar{g}_{ij}| / \sigma_{ij} \tag{1}$$

where $\bar{\mathbf{g}}_i$ is the vector of mean intensities for the region and σ_{ij} is an estimate of the mean absolute difference from the mean across a training set.[1]

We can then search new images with such a model, by performing an exhaustive search at a range of positions, orientations and scales to locate local minima of $f_i(\mathbf{g}(\mathbf{t}))$. This result in multiple responses for each patch [1].

2.2 Geometric Relationships

To disambiguate the multiple responses of a single patch model, we create a model containing a set of N patch models, together with a model of the pairwise relationships between them. This is a widely used and effective technique [6].

Given multiple possible candidates for each part position (from the patch detectors), we used a graph algorithms to locate the optimal solutions. We used a variant of dynamic programming in which a network is created where each node can be thought of as having at most two parents. Details of this method are discussed in [1].

Each candidate response for part i has a pose with parameters $\mathbf{t}_i = \{\mathbf{p}_i, s_i, \theta_i\}$. The relationship between part i and part j can be represented in the cost function, $f_{ij}(\mathbf{t}_i, \mathbf{t}_j)$. This can be derived from the joint PDF of the parameters.

[1] We find this form (which assumes the data has an exponential distribution) gives more robust results than normalised correlation, which is essentially a sum of squares measure.

In the following we take advantage of the fact that the orientation and scale of the objects are approximately equivalent in each image, and simply use a cost function based on the relative position of the points:

$$f_{ij}(\mathbf{t}_i,\mathbf{t}_j) = ((\mathbf{p}_j - \mathbf{p}_i) - \mathbf{d}_{ij})^T \mathbf{S}_{ij}^{-1} ((\mathbf{p}_j - \mathbf{p}_i) - \mathbf{d}_{ij}) \quad (2)$$

where $\mathbf{d}_{ij}$ is the mean separation of the two points, and $\mathbf{S}_{ij}$ is an estimate of the covariance matrix.

The matching algorithm thus seeks to find the candidates which minimise the following function

$$F = \sum_{i=1}^{N} f_i(\mathbf{g}_i) + \alpha \sum_{(i,j)\in Arcs} f_{ij}(\mathbf{p}_i,\mathbf{p}_j) \quad (3)$$

The value of α affects the relative importance of patch and geometry matches. In the following we use $\alpha = 0.1$, chosen by preliminary experiments on a small subset of the data. Ways of automatically choosing a good value of α are the focus of current research.

2.3 Building the Model

We initialise a model using a set of parts defined by boxes placed on a single image by the user (for instance, the rectangles shown in Figure 1a). This takes about one minute to do, and allows the algorithm to take advantage of user supplied knowledge. We then automatically define a set of connecting arcs based on the distances between the centres of the boxes. We use a variant of Prim's algorithm for the minimum spanning tree, where each node has two parent nodes, rather than one [1].

We then refine the model by applying it to the whole dataset, ranking the results by final fit value (per image), and building statistical models of intensity and pairwise relationship from the best 50% of the matches.

2.4 Dense Correspondence

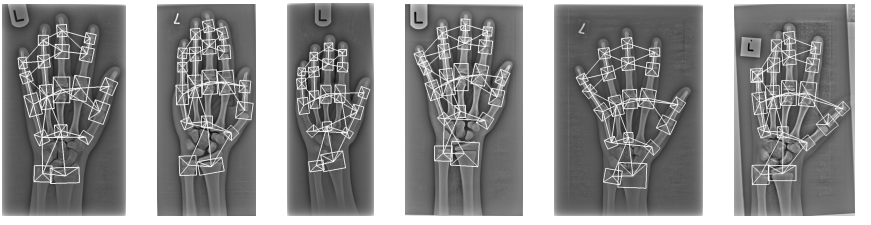
At convergence we obtain a model of parts and geometry, together with a sparse annotation of every image in the training set. The centres of each part region define correspondences.We use these to initialise a groupwise registration. We place a dense mesh of control points on the first image, use a thin-plate spline based on the sparse annotation to propagate these points to all other images. We then compute the mean shape and warp each example into the mean. Furthermore we perform non-rigid registration [3] to modify the control points on each image to best match to the mean. Finally we re-compute the mean and iterate.

3 Experiments

We applied the technique described above to a set of 536 radiographs of the hands of children, taken as part of another study[2]. We divided the dataset into three age-groups. AgeGroup1 -63 images (5 - 7 yrs), AgeGroup2 -284 images (8-13 yrs) and AgeGroup3 - 189 images (14 -19 years) In our earlier work [1] we found the optimal number of boxes to be 19 boxes. These 19 boxes were annotated on one image (see Figures 1a). For each choice of boxes on

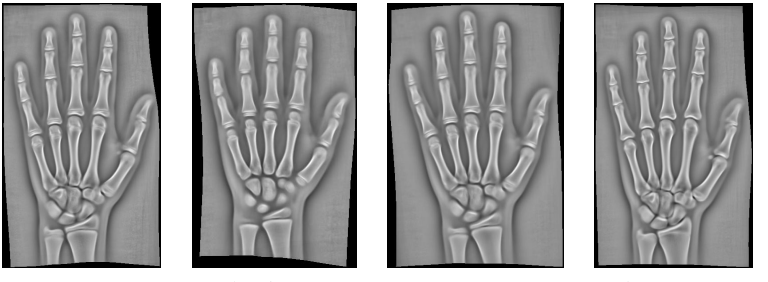
[2]The authors would like to thank K.Ward, R.Ashby, Z. Mughal and Prof.J.Adams for providing the images.

a single image, a model of parts and geometry was constructed and used to locate equivalent points on other images. The models were then rebuilt and refined as described above. Figure 1a shows the initial 19 boxes on one of the images, together with the automatically chosen connectivity. Matches with the final model are shown in Figure 1b,c,d,e for the various groups and an example of failure in 1f. The found points in each of the groups were used to initialise a groupwise algorithm as described above. Qualitative results of the registration is shown in Figures 2. The crispness of the images indicate a good alignment.



a) b) c) d) e) f)

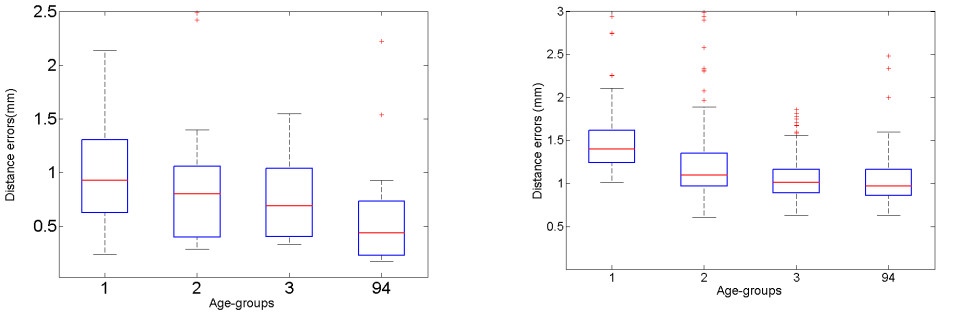
Figure 1: Example of model(a), search results with 19 parts for set94(b) [], AgeGroup1(c), AgeGroup2 (d), AgeGroup3 (e) and an example of a failure (f) respectively (see the tip of the fifth finger near the label).



a) set94 b) 5-7yr c) 8-13yr d) 14-19yr

Figure 2: Final mean images after groupwise registration. a) set94 [], b) AgeGroup1, c)AgeGroup2 and c) AgeGroup3.

We evaluated the accuracy of the points location by comparing with manual annotations based on an evaluation framework formulated in []. The mean distance errors for sparse point errors was found to be $0.70 \pm 0.08mm$, $1.08 \pm 0.18mm$, $0.91 \pm 0.15mm$, $0.75 \pm 0.09mm$ for the set94 (images used in []), AgeGroup1, AgeGroup2, AgeGroup3 respectively. The result of AgeGroup3 14 -19, a very difficult group, is comparable to the original result obtained in []. Figure 3a presents the distribution of the errors and compare the various groups. For the dense correspondence accuracy, a median error of $0.94mm$, $1.38mm$, $1.1mm$ and $1.01mm$ for the set94, AgeGroup1, AgeGroup2, AgeGroup3 respectively. These errors are higher than in sparse point placement because the evaluation is based on the entire image region []. Figure 3b presents the distribution of the errors and compare the various groups. Note that in both cases errors are highest for AgeGroup1. The few number of images and very large variation may be responsible. Sometimes there is no correspondence amongst the bones.



a) b)

Figure 3: Comparison of statistics of points errors for various groups. a) Accuracy of sparse point placement and b) Errors after groupwise registration (mm).

4 Discussion and Conclusions

We have evaluated an approach for automatically locating sparse correspondences across a set of images, by constructing a parts and geometry model with an extended dataset. We achieve an accuracy of 0.75mm on the positioning of the chosen parts. This is significantly better than results quoted by Donner *et al.*[5] (approx. 1.5mm, though on a different, smaller dataset). The found points are sufficient to initialise a more detailed group-wise registration which can give dense point correspondences with approximately 1mm accuracy over the whole hand. We can conclude that these results are comparable with our earlier work [1]. We have commenced more work on the AgeGroup1 to achieve higher accuracy.

References

[1] Steve A. Adeshina and Timothy F. Cootes. Constructing part-based models for groupwise registration. In *Proc. IEEE International Syposium on Biomedical Imaging*, 2010.

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

[3] T.F. Cootes, C.J. Twining, V.Petrović, R.Schestowitz, and C.J. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In *16th British Machine Vision Conference*, volume 2, pages 879–888, 2005.

[4] R.H. Davies, C.Twining, T.F. Cootes, J.C. Waterton, and C.J. Taylor. 3D statistical shape models using direct optimisation of description length. In *European Conference on Computer Vision*, volume 3, pages 3–20. Springer, 2002.

[5] Rene Donner, Horst Wildenauer, Horst Bischof, and Georg Langs. Weakly supervised group-wise model learning based on discrete optimization. In *Proc. MICCAI*, volume 2, pages 860–868, 2009.

[6] P.F. Felzenszwalb and D.P.Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79, 2005.