Automatic Learning Sparse Correspondences for Initialising Groupwise Registration

Pei Zhang, Steve A. Adeshina, and Timothy F. Cootes

Imaging Science and Biomedical Engineering, The University of Manchester, UK
{Pei.Zhang-2,steve.adeshina}@postgrad.manchester.ac.uk,
tim.cootes@manchester.ac.uk

Abstract. We seek to automatically establish dense correspondences across groups of images. Existing non-rigid registration methods usually involve local optimisation and thus require accurate initialisation. It is difficult to obtain such initialisation for images of complex structures, especially those with many self-similar parts. In this paper we show that satisfactory initialisation for such images can be found by a parts+geometry model. We use a population based optimisation strategy to select the best parts from a large pool of candidates. The best matches of the optimal model are used to initialise a groupwise registration algorithm, leading to dense, accurate results. We demonstrate the efficacy of the approach on two challenging datasets, and report on a detailed quantitative evaluation of its performance.

1 Introduction

Groupwise non-rigid image registration is a powerful tool to automatically establish dense correspondences across large sets of images of similar but varying objects. Such correspondences are widely used, for instance to construct statistical models of shape or appearance [1]. These models have a wide range of applications in medical image processing, such as segmentation of anatomical structures or morphometric analysis.

Existing groupwise techniques [2,3,4] generally treat registration as an optimisation problem which is solved with local minimisation methods. As such they are sensitive to initialisation and will fail if "good-enough" starting points are not found. This problem is prominent when registering images of objects with considerable shape variation and multiple similar sub-structures, such as radiographs of the human hand (Fig. 1a). The reason is that non-rigid registration methods usually use an affine transformation to find an approximate initialisation, then refine this with local optimisation. Unfortunately this is insufficient for objects with complex structures because there are too many local minima. For instance, Fig. 1b shows the average of a set of hand radiographs after affine alignment. This shows that the registration failed to register the fingers adequately on many of the examples.

Where objects have a number of distinctive parts, these can be individually located and used to initialise further registration. However, important classes of

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 635–642, 2010.

a) Hand Examples b)Affine mean c) Spine examples

Fig. 1. Examples of hand and spine radiographs used in experiments

object (such as the human hand or spine) contain multiple repeating structures, leading to inherent ambiguity when attempting to match any given part. An effective way to resolve this ambiguity is to use a geometric model of a set of parts, that is, the parts+geometry model [5,6]. Such models are often created so as to optimise their ability to detect or discriminate objects [7]. In this paper we optimise them so as to achieve accurate correspondence.

In [8] we have shown that careful selection of a small number of parts on a single image can be used to construct a parts+geometry model from the whole set, which can lead to good sparse correspondences. These can be used to initialise a dense registration, leading to accurate results. In this paper we describe an algorithm which avoids the manual intervention, and automatically selects a good set of parts for a sparse model, so that the whole process of correspondence establishment is still fully automatic.

The work most similar to ours is that of Langs et al. [9,10]. They describe a method of constructing sparse shape models from unlabelled images, by finding multiple interest points and using minimum description length (MDL) principle to determine optimal correspondences, finding the model which minimises the description of the feature points. Another related approach was developed by Karlsson and Åström [11], who built patch models to minimise an MDL function, estimating the cost of explaining the whole of each image using the patches (by including a cost for the regions not covered by patches).

Both of the above approaches represent shape with a Point Distribution Model [1]. Such representations are useful for local optimisation, but cannot efficiently deal with multiple candidates. By instead learning a parts+geometry model, where the geometry is modelled with a sparse graph, we can take advantage of dynamic programming (DP) algorithms which can efficiently find the global optima where multiple candidates are present. Like Karlsson and Åström, our cost function is based on explaining the whole of an image region, but in our case this is done by constructing a model of the whole image using non-rigid deformation based on the centres of the part models. In addition our goal is somewhat different—we seek a sparse set of parts which can be used to initialise a local optimisation based groupwise registration scheme.

Our main contributions are **1)** automatic selection of a suitable subset of feature parts which are likely to be well localised; **2)** unsupervised learning

of a parts+geometry model which can be used to obtain the optimal sparse correspondences; **3)** a comparison of two different methods of locating the parts.

2 Part Models

We first construct a set of candidate parts which represent structures that are present in most of the images. Given a set of such parts we can automatically construct a geometric model of their relative positions (see Sect. 3). The aim then is to select a subset of parts which leads to the model that determines the best set of dense correspondences between the training images.

We describe the position $\mathbf{x} = (x, y)$, scale s and orientation θ of each part with a pose parameter $\mathbf{p} = \{\mathbf{x}, s, \theta\}$. We have experimented with two methods of representing and localising parts—patch based and SIFT based:

Patch Models: Each part is represented as a statistical model of the intensities over an oriented square region centred at $\mathbf{x}$. Let $\mathbf{g}(I, \mathbf{p})$ be the intensities sampled from the region defined on image I with the pose parameter $\mathbf{p}$, normalised to have a mean of zero and unit variance. The quality of fit to such a model is evaluated as $f_i(\mathbf{g}(I, \mathbf{p})) = \beta \sum_{j=1}^{n} |g_j - \bar{g}_{ij}|/\sigma_{ij}$, 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. β is a normalisation factor chosen so that the standard deviation of the best fits across the training set is unity[1].

Locating candidates for each part involves a multi-resolution search at a range of scales and orientations, in which local optima are located at a coarse scale and refined at finer scales. This approach allows us to quickly search large regions, usually resulting in a few tens of hypotheses.

In order to obtain a set of part models for an unlabelled image set, we arbitrarily choose one image as the reference image[2]. Then we use it to generate a group of patches for a range of sizes, arranged in an overlapping grid pattern (Fig. 2a). We use the region within a given patch to build a part model, and search the rest of the images for the best match on each. We rank the best matches by the quality of fit, and rebuild the model from the best 50% of these. The resulting model is then used to search the images again to find its matches.

To select the models which are likely to have good localisation ability from the set, we sort them by how well the optima is localised [3], and select the best. The set of retained models is denoted as $\mathbf{R}_{\text{sub}}$. Figure 2b shows some examples of the selected part models.

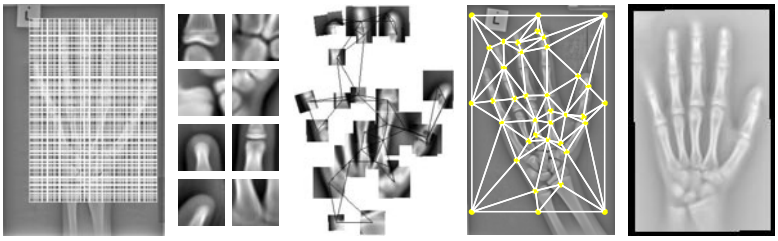
SIFT Models: We can also represent each part using a SIFT signature[4] [12]. To search for candidates for a part we compute interest points using **1)** an edge

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

[2] we find that the algorithm is insensitive to the choice of the reference image.

[3] This is evaluated by $q_i = \min_{\delta\mathbf{x}} \|\bar{\mathbf{g}}(I, \mathbf{p}_i + \delta\mathbf{x}) - \bar{\mathbf{g}}(I, \mathbf{p}_i)\|_2$, where $\bar{\mathbf{g}}(I, \mathbf{p}_i) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{g}(I_k, \mathbf{p}_{i,k})$, $\mathbf{p}_{i,k}$ is the pose of the best match of the part model i on image I_k, $\delta\mathbf{x}$ is the displacement of the model ($1 \leq |\delta\mathbf{x}| \leq 4$ in this paper). If q is small then the model is not good at localising.

[4] `http://www.vlfeat.org`



(a) (b) (c) (d) (e)

Fig. 2. (a) An overlapping grid. (b) Typical patch models. (c) A parts+geometry model. (d) Sparse points used for correspondence. (e) The resulting mean reference.

detector and **2)** a variant of a local symmetry point detector [13], which returns local minima of a smoothed edge strength image. We then rank the points using their matches to the part signature and retain the best 50.

To generate a set of initial candidate part models, we locate interest points on every image. We then select those on the reference image which pass a variant of forward-backward matching [14]. A point is selected and added to $\mathbf{R}_{\text{sub}}$ if for at least 50% of the rest of the images in the set, the best match in the other image also matches back to the original point.

3 Parts+Geometry Models

Model Definition: Given a set of m parts from $\mathbf{R}_{\text{sub}}$, we can automatically build a parts+geometry model $\mathcal{G}$ (see below), which can be used to disambiguate the multiple responses of a single part model. An example is given in Fig. 2c. $\mathcal{G}$ represents the object of interest using the parts together with a collection of pairwise geometric relationships between each part, which are defined by a set of arcs ω. For each part, we take its best K matches on each image as its match candidates.

Let $f_i(\mathbf{p})$ be the cost associated with part i having pose $\mathbf{p}$. The relationship between part i and part j can be represented in the cost function, $f_{ij}(\mathbf{p}_i, \mathbf{p}_j)$. This can be derived from the joint pdf of the parameters. In the following we take advantage of the fact that the orientation and scale of the objects are roughly equivalent in each image, and simply use a cost function based on the relative position of the parts, $f_{ij}(\mathbf{p}_i, \mathbf{p}_j) = ((\mathbf{x}_j - \mathbf{x}_i) - \mathbf{d}_{ij})^{\mathrm{T}} \mathbf{S}_{ij}^{-1} ((\mathbf{x}_j - \mathbf{x}_i) - \mathbf{d}_{ij})$, where $\mathbf{d}_{ij}$ is the mean separation of the two parts, and $\mathbf{S}_{ij}$ is an estimate of the covariance matrix. If the objects are likely to undergo significant scaling or orientation changes across the set, more sophisticated function can be used.

To find the best match of $\mathcal{G}$ on an image, we must select one match candidate point $\mathbf{p}_i$ for each part (from its K match candidates) so as to minimise the following function

$$C = \sum_{i=1}^{m} f_i(\mathbf{p}_i) + \alpha \sum_{(i,j)\in\omega} f_{ij}(\mathbf{p}_i, \mathbf{p}_j). \tag{1}$$

The value of α affects the relative importance of part and geometry matches. Given multiple possible candidates for each part, we can use graph algorithms to locate the optimal solutions to (1). We use a method which is similar to that used in [15], in which a network is created where each node can be thought of as having at most two parents. The optimal solution for this can be obtained with a variant of DP algorithm, in $O(mK^3)$ time. If K is modest, this is still fast.

Model Construction: Given a set of parts in the reference image, we can construct a set of connecting arcs, ω, in such a way that each point other than the first two is connected to two parents [8]. The geometric relationships for each arc $(i, j) \in \omega$ are initialised with Gaussians (with standard deviation set to 25% of the length of the arc in the reference image). We then refine the model by applying it to the responses on each image, ranking the results by final fit value (per image), and re-estimating the geometric distributions from the results on the best 50% of the images—essentially a form of robust model building.

Model Evaluation: The match of $\mathcal{G}$ to each image defines a set of sparse correspondences. To evaluate the performance of $\mathcal{G}$ we estimate how effectively a model built from its matches can represent the original image set (a description length). To calculate this we augment the points from each best match and with a set of fixed border points on each image (Fig. 2d). We then align the sets of points and compute the mean. We create a triangulation of the mean and use it to warp each image into the reference frame. We compute the mean intensity in the reference frame (Fig. 2e). Finally, we warp the reference image into the frame of each target image, and compute the sum of absolute differences over the region of interest $U(\mathcal{G}) = \sum_{k=1}^{N-1} \sum_{\mathbf{x} \in R} \left\| I_k(\mathbf{x}) - I_0(W^{-1}(\mathbf{x})) \right\|$, where $I_k(\mathbf{x})$ is the target image intensity at $\mathbf{x}$, R is the region of interest in the target frame and $W(\mathbf{y})$ is the transformation from reference I_0 to target I_k.

Model Selection: Each subset of parts from $\mathbf{R}_{\text{sub}}$ leads to a model, whose quality can be measured as U. Selecting the best subset of parts is thus a combinatoric problem. We solve it using a population based optimisation algorithm, which is similar to the Genetic Algorithm, to find the $\mathcal{G}$ with the minimal U. We create an initial population by randomly sampling subsets from $\mathbf{R}_{\text{sub}}$. For each set we generate a $\mathcal{G}$ then evaluate it. We then rank the members of the population (each a candidate set of parts) by U. We discard the worst 50%, and generate new candidates from pairs of candidates randomly selected from the best 50%. To generate a new subset we simply randomly sample from the union of parts from the two candidate parent sets. Repeating the above process leads to the best $\mathcal{G}$, whose best matches define the optimal set of sparse correspondences.

4 Establishing Dense Correspondences

To obtain an accurate dense registration, we use the sparse correspondences to initialise a non-rigid registration algorithm [4]. We use a coarse to fine algorithm to improve efficiency and robustness. The approach is to compute a robust mean (Fig. 3b) using the current estimate of the deformation field, then to refine the correspondences to improve the match to this mean (Fig. 3c).

5 Experiments

To demonstrate the approach, we applied it to two different datasets: **1)** 94 radiographs of the hands of children (aged between 11-13), taken as part of study into bone ageing (Fig. 1a). Each image has been marked with 37 points by a human expert; **2)** 106 radiographs of the lumbar spine (Fig. 1b). Each image has 337 manual landmarks placed around the outline of the vertebrae. Both sets of images have a resolution of 0.2mm per pixel and a height of about 1300 pixels.

By systematically initialising patch models in an overlapping grid on the reference image at a range of sizes, we automatically constructed over 1900 and 500 part models for the hand and spine sets respectively. We ranked by their localisability and selected the best 500 for the hands and the best 100 for the spines. We used the population based optimisation to select the best parts+geometry models from the two sets, for a range of different numbers of parts. An example of the resulting parts+geometry models for the hand set is shown in Fig. 3a. The resulting correspondences were used to estimate an initial dense correspondence, from which a robust mean was estimated (Fig. 3b). Groupwise non-rigid registration was applied, giving the final result shown in Fig. 3c. Equivalent results for the spine set are shown in Fig. 3d-f.

We also repeated the experiment on the hands using the SIFT based models.

To evaluate the accuracy of the result we compare with a manual annotation. We used the resulting dense correspondences to warp each set of manual landmarks into a reference frame, computing their mean. We then projected this mean back to each individual image (Fig. 4) and calculated the mean absolute difference between each mean point and the original manual annotation. For the spines, we calculated the mean distance between the mean points and the curve formed by the original manual landmarks.

Table 1 shows the statistics of the resulting point location errors for different numbers of parts for the two datasets. There is not a clear relationship between the number of parts and performance—once sufficient parts are available to cover the main components of the object, adding further elements is unlikely to improve performance and may lead to a decline. We found that the patch based

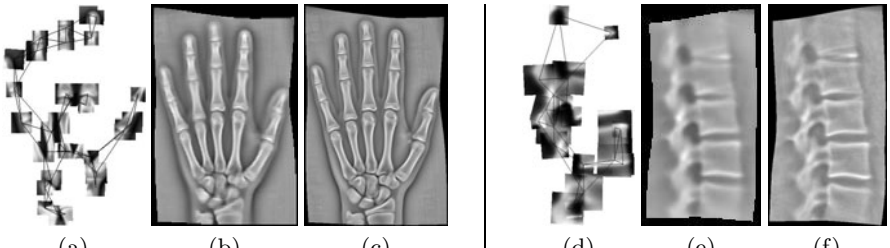


Fig. 3. Examples of the best parts+geometry models and registered means

Table 1. Point location errors (mm) of the dense correspondence

a) Hand set

Parts	SIFT based			Patch based		
	Mean	Med.	90%	Mean	Med.	90%
5	4.2	3.6	6.5	3.3	2.4	7.8
10	2.0	1.2	3.7	2.2	1.7	4.0
15	1.7	1.1	2.9	1.1	1.0	1.6
20	1.4	1.1	2.3	1.3	0.9	2.5
25	1.1	0.9	1.9	1.1	0.9	1.7
30	1.6	1.3	2.9	**1.0**	**0.8**	**1.4**

b) Spine set (patch based only)

Parts	Mean	Med.	90%
5	6.8	3.1	21.9
10	5.2	2.5	13.2
15	5.0	3.1	9.7
20	6.2	3.9	14.5

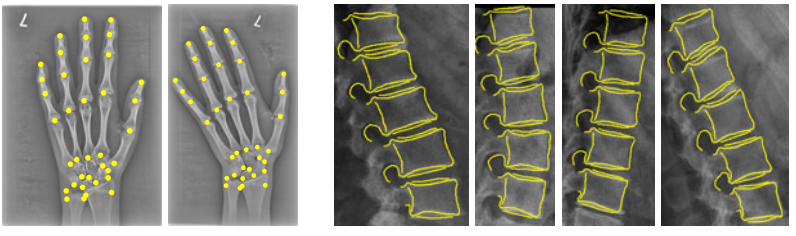


Fig. 4. Examples of projection of average points onto individual image

part models generally lead to better results than their SIFT based counterparts, though locating them is more computationally expensive.

Computing the equivalent error for the hands after performing a standard affine initialisation gives a median error of 12.0mm, which is little improved by further dense registration due to local minima. The parts+geometry method gives substantially better results, demonstrating the importance of good initialisation. Comparison with other published approaches is not easy. The most similar approach in the literature [10] evaluates on a set of only 20 hand radigraphs (0.34mm/pixel), obtaining a mean error of 2.0mm and a median of 1.7mm (though on a larger set of points)—our method appears to give significantly better results.

6 Conclusions and Future Work

We have described an approach for automatically locating dense correspondences across a set of images, by using the sparse matches of a parts+geometry model to initialise groupwise non-rigid registration. It is able to achieve good results on two challenging datasets. The technique potentially can be used for a wide range of other datasets.

In the above we show results for particular choices of numbers of parts. Potentially it will be possible to use the framework to estimate the optimal number of parts (as is done by [11]). However, our objective is not to explain the whole image with the parts+geometry model, just to obtain good enough correspondence

for the dense registration. The most efficient method of selecting the appropriate number is currently under investigation. Moreover, we have used a relatively simple topology for the geometric model in order to allow efficient global optimisation. In future we will try using more highly connected graphs.

References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
2. Baker, S., Matthews, I., Schneider, J.: Automatic construction of active appearance models as an image coding problem. IEEE TPAMI 26(10), 1380–1384 (2004)
3. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)
4. Cootes, T.F., Twining, C.J., Petrović, V., Schestowitz, R., Taylor, C.J.: Groupwise construction of appearance models using piece-wise affine deformations. In: Proc. of British Machine Vision Conference, vol. 2, pp. 879–888 (2005)
5. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. International Journal of Computer Vision 61(1), 55–79 (2005)
6. Donner, R., Micusik, B., Langs, G., Szumilas, L., Peloschek, P., Friedrich, K., Bischof, H.: Object localization based on markov random fields and symmetry interest points. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 460–468. Springer, Heidelberg (2007)
7. Fergus, R., Perona, P., Zisserman, A.: Weakly supervised scale-invariant learning of models for visual recognition. International Journal of Computer Vision 71, 273–303 (2007)
8. Adeshina, S.A., Cootes, T.F.: Constructing part-based models for groupwise registration. In: Proc. of International Symposium on Biomedical Imaging (2010)
9. Langs, G., Peloschek, P., Donner, R., Bischof, H.: Annotation propagation by MDL based correspondences. In: Proc. of Computer Vision Winter Workshop (2006)
10. Langs, G., Donner, R., Peloschek, P., Bischof, H.: Robust autonomous model learning from 2D and 3D data sets. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 968–976. Springer, Heidelberg (2007)
11. Karlsson, J., Åström, K.: MDL patch correspondence on unlabeled images with occlusions. In: Proc. of Computer Vision and Pattern Recogition (2008)
12. Lowe, D.: Object recognition from scale invariant features. In: Proc. of International Conference on Computer Vision, vol. 2, pp. 1150–1157 (1999)
13. Donner, R., Micusik, B., Langs, G., Bischof, H.: Sparse MRF appearance models for fast anatomical structure localisation. In: Proc. of British Machine Vision Conference, vol. 2, pp. 1080–1089 (2007)
14. Fua, P.: A parallel stereo algorithm that produces dense depth maps and preserves image features. Machine Vision and Applications 6, 35–49 (1993)
15. Felzenszwalb, P.F., Huttenlocher, D.P.: Representation and detection of deformable shapes. IEEE TPAMI 27(2), 208–220 (2005)