Initialising Groupwise Non-rigid Registration Using Multiple Parts+Geometry Models

Pei Zhang[1], Pew-Thian Yap[1], Dinggang Shen[1], and Timothy F. Cootes[2]

[1] Department of Radiology and Biomedical Research Imaging Center (BRIC)
The University of North Carolina at Chapel Hill, USA
`peizhang@email.unc.edu`, {`ptyap,dgshen`}`@med.unc.edu`
[2] Imaging Sciences, School of Cancer and Enabling Sciences,
The University of Manchester, UK
`timothy.f.cootes@manchester.ac.uk`

Abstract. Groupwise non-rigid registration is an important technique in medical image analysis. Recent studies show that its accuracy can be greatly improved by explicitly providing good initialisation. This is achieved by seeking a sparse correspondence using a parts+geometry model. In this paper we show that a single parts+geometry model is unlikely to establish consistent sparse correspondence for complex objects, and that better initialisation can be achieved using a set of models. We describe how to combine the strengths of multiple models, and demonstrate that the method gives state-of-the-art performance on three datasets, with the most significant improvement on the most challenging.

1 Introduction

Finding consistent correspondences across sets of images is a challenging problem, with applications in many areas, particularly in constructing statistical models of shape or appearance [1,2]. A promising solution is groupwise non-rigid image registration [3,4,5], where the correspondence is usually determined by minimising some objective function. A common choice of initialisation for groupwise registration is affine transformation. However, our recent studies have shown that it is insufficient when registering images of large local shape variations and repeating structures [6,7]. Hence, more sophisticated approaches to initialisation have been developed [8,9,6,7].

Although implementation varies, these algorithms share the same idea, that is, deliberately finding a sparse set of corresponding points and using them to initialise groupwise registration. In [9], it was shown that a small set of manually selected parts can be used to build a parts+geometry model [10,11,12,13] capable of giving a good sparse correspondence. Later in [6,7], such human intervention was shown to be unnecessary—the set of "good" parts can be automatically obtained. Instead of using parts+geometry models, Langs et al. [8] explored using a Point Distribution Model [1] in a weakly supervised fashion. The model is iteratively estimated so as to minimise the description length of the feature points across the images, leading to the optimal correspondence.

A problem with the above methods is that they all attempt to use a *single* model to achieve the good sparse correspondence. Due to the limited amount of sample images and the imperfection of the algorithm, it is unlikely for a single model to capture every

(a) Model A (b) Consistent matches (c) Failures

(d) Model B (e) Failures (f) Consistent matches
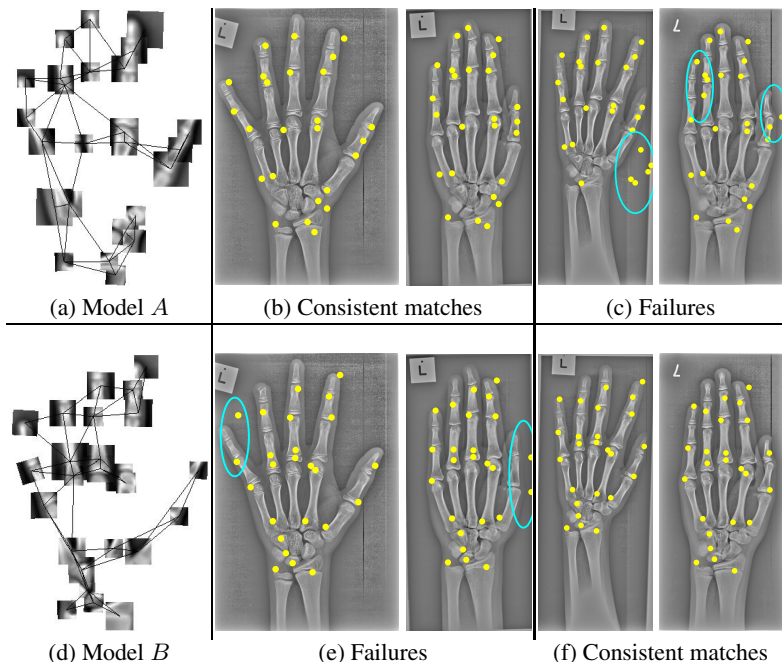
Fig. 1. Left column: two parts+geometry models. Right four columns: matches of the models. Top row: model A finds consistent matches on two images but fails on the other two. Bottom row: model B gives consistent matches on those where model A fails, but fails where model A works well. Failures are indicated by cyan ellipses.

possible variation of the object, particularly those of complex structures. As a result, the learned model may fail on some of the images. See Fig. 1 for an example.

Figure 1 also suggests that it is always possible to learn a model that can deal well with a subset of images. Different models that work well with different subsets are likely to complement each other. By combining the best result from each model, we may achieve better initialisation than just using a single model.

In this paper we explore using multiple parts+geometry models to initialise groupwise non-rigid registration—a *multi-model* initialisation scheme. We use our previous approach [6] to generate a population of reasonably good models. We then apply the algorithm described below to these models to obtain the desired correspondence.

2 Initialisation with Multiple Parts+Geometry Models

Let the term *pattern* refer to the sparse set of points found on an image by a particular model, which thus defines the correspondence for that image to a reference frame for that model. When using multiple models we obtain multiple patterns on each image. As shown in Fig. 1, some patterns are apparently better than the others. By replacing those poor patterns with good ones, we can modify and improve the correspondence defined by the single model.
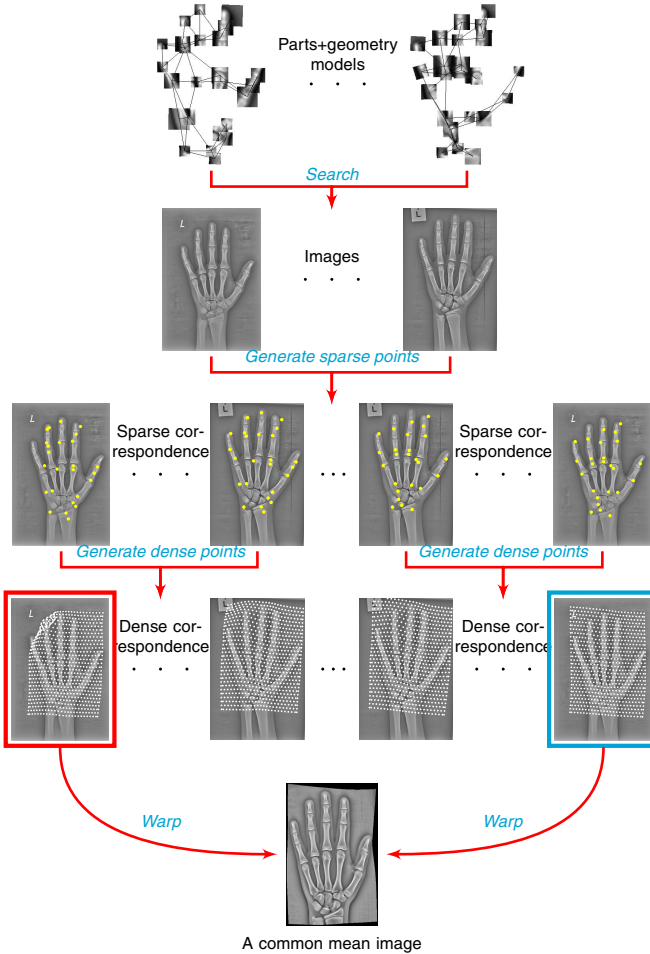
Fig. 2. An overview of the multi-model initialisation strategy

An overview of the multi-model initialisation scheme is given in Fig. 2. Given a set of parts+geometry models, we use each in turn to search the image set to obtain a sparse correspondence, where a dense correspondence can be generated using a thin-plate spline (TPS) interpolation. Poor patterns will lead to poor dense points (indicated by the red box in Fig. 2). The quality of the pattern can be evaluated by warping the target image to a mean image using the dense points, and comparing the similarity between the warped image and the mean. To choose the best pattern for an image, we use each of the associated sets of dense points to warp the image to the mean and compute the similarity. A new correspondence can thus be established by grouping the sets of dense points related to the best patterns across the images. We give the details of each step below.

2.1   Dense Points

To easily compare the quality of different patterns for an image, we transfer the corre-spondence information encoded in each pattern to a common set of points. A simple method is to generate a dense set of points on a reference image and propagate the points to the other images using a TPS. We use $\mathbf{X} = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)^{\mathrm{T}}$ to denote the positions of the dense points.

2.2   Quality of the Pattern

An observation from groupwise registration is that if the correspondence across the im-age set is well established, we should be able to obtain a crisp mean image. Furthermore, if a pattern is good, its related image should be similar to the mean when comparing the two in the same frame, and *vice versa*. Hence, we can use a mean image to evaluate the quality of a pattern.

Suppose we have a set of images $\{I_k | k = 1, \ldots, N_I\}$ and N_G parts+geometry mod-els. We use $\{\mathbf{X}_k^l | l = 1, \ldots, N_G\}$ to denote the sets of dense points associated with image I_k. Let $\bar{I}$ be the mean image and $\overline{\mathbf{X}}$ be the dense points in the mean (see below). Given an image I_k and one of its dense point sets $\mathbf{X}_k^l$, a warp from $\bar{I}$ to I_k is uniquely defined by $\overline{\mathbf{X}}$ and $\mathbf{X}_k^l$. We write this warp as $\mathbf{z}' = W(\mathbf{z} : \overline{\mathbf{X}}; \mathbf{X}_k^l)$, where $\mathbf{z}$ is a point in $\bar{I}$ and $\mathbf{z}'$ is the corresponding point in I_k. To evaluate the quality of the pattern related to $\mathbf{X}_k^l$, we warp I_k onto $\bar{I}$ so as to compare them in a same frame, and use the following function

$$D_k^l = \sum_{\mathbf{z} \in R} \left| I_k(W(\mathbf{z} : \overline{\mathbf{X}}; \mathbf{X}_k^l)) - \bar{I}(\mathbf{z}) \right|, \tag{1}$$

where R is a region of interest in the mean frame. This function computes the absolute intensity difference over the region of interest between the warped image $I_k(W(\mathbf{z} : \overline{\mathbf{X}}; \mathbf{X}_k^l))$ and the mean. Note that all the images have been preprocessed to standardise their intensity ranges.

2.3   Mean Image

Given $\{\mathbf{X}_k^l | k = 1, \ldots, N_I\}$, we can perform the following steps to compute the mean image $\bar{I}^l$ and its associated dense points $\overline{\mathbf{X}}^l$:

(1) Align each $\mathbf{X}_k^l$ to a reference frame[1] using Procrustes Analysis [1]. Averaging the aligned dense sets of points leads to $\overline{\mathbf{X}}^l$;
(2) Create a triangulation of $\overline{\mathbf{X}}^l$ using the Delaunay algorithm;
(3) Warp each image I_k to the reference frame, computing $I_k(W(\mathbf{z} : \overline{\mathbf{X}}^l; \mathbf{X}_k^l))$,[2] $\mathbf{z} \in R$;
(4) Compute the mean image using $\bar{I}^l(\mathbf{z}) = \frac{1}{N_I} \sum_{k=1}^{N_I} I_k(W(\mathbf{z} : \overline{\mathbf{X}}^l; \mathbf{X}_k^l)), \mathbf{z} \in R$.

---

[1] The choice of the reference frame is free. Any image in the set can be used for this purpose.
[2] $I_k(W(\mathbf{z} : \overline{\mathbf{X}}^l; \mathbf{X}_k^l))$ is computed by piece-wise linear interpolation between corresponding triangles in $\overline{\mathbf{X}}^l$ and $\mathbf{X}_k^l$. We use piece-wise linear interpolation for efficiency.

A Common Mean. Different sets of points $\{\mathbf{X}_k^l | k = 1, \ldots, N_I\}$ will lead to a different mean image $\bar{I}^l$. Hence, the quality D computed using different means cannot be compared directly. This can be solved by using a single, common mean. We take the following steps to compute this common mean image:

(1) For $l \leftarrow 1$ to N_G
- (a) Compute the mean image $\bar{I}^l$ using the sets of points $\{\mathbf{X}_k^l | k = 1, \ldots, N_I\}$;
- (b) Warp every image to the mean $\bar{I}^l$, computing D_k^l using (1);
- (c) Rank all images by D_k^l and select the top 50% of the images as the set $\mathcal{S}^l$;

(2) Find a common set of images $\mathcal{S} = \bigcap_{l=1}^{N_G} \mathcal{S}^l$—selecting the images in all sets $\{\mathcal{S}^l\}$;

(3) For each image in $\mathcal{S}$, average associated sets of dense points, $\widehat{\mathbf{X}}_k = \frac{1}{N_G} \sum_{l=1}^{N_G} \mathbf{X}_k^l$;

(4) Compute a mean image $\bar{I}$ using $\mathcal{S}$ and the sets of points $\{\widehat{\mathbf{X}}_k\}$ (similar to what has been described above). Take this as the common mean.

2.4 Pattern Selection

Once we have the common mean image, we can take the following steps to select the best pattern for each image:

(1) For $l \leftarrow 1$ to N_G
  For $k \leftarrow 1$ to N_I
    Warp I_k to the common mean and compute D_k^l;

(2) For $k \leftarrow 1$ to N_I
  Select the best $\mathbf{X}_k^{\hat{l}}$ from $\{\mathbf{X}_k^l | l = 1, \ldots, N_G\}$ such that $D_k^{\hat{l}}$ is minimum.

By grouping $\mathbf{X}_k^{\hat{l}}$ across the image set we can obtain a new correspondence $\{\mathbf{X}_k^{\hat{l}} | k = 1, \ldots, N_I\}$. Although we can use it to directly initialise groupwise registration, in the following experiment we use it to generate a sparse correspondence. This is more efficient, and allows fairer comparison with the single-model scheme which only outputs sparse sets of points.

To create the sparse correspondence we (1) generate a sparse set of points on an image I_k using the best parts+geometry model (in terms of model utility, see [6] for details); (2) project the sparse points onto the other images using the piece-wise affine transformation between different $\mathbf{X}_k^{\hat{l}}$.

3   Experiments

We demonstrate the approach on three different datasets of increasing difficulty:

(1) 100 digital micrographs of female fly wings. Each image has a size of 1280×1022 pixels and is marked with 15 points by human expert (Fig. 3a);

(2) 100 radiographs of the hands of children (aged between 10-13), taken as part of study into bone ageing. The image size varies across the set, with a height ranging from 1000 to 1700 pixels. Each image has 37 manually labelled landmarks. The resolution of this set of images is 0.2mm per pixel (Fig. 3b);

(a) Fly wing (b) Hand (c) Spine
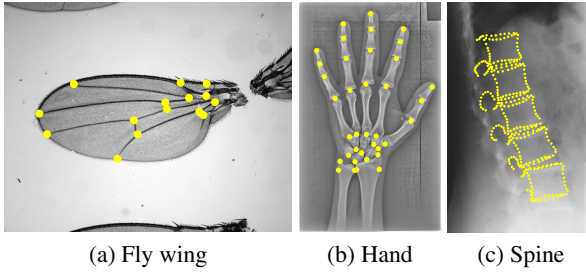
Fig. 3. Examples of the datasets and associated landmarks used in the experiment

(3) 100 radiographs of the lumber spine. The image height varies from 1500 to 1700 pixels. Each image has 337 manual landmarks placed around the outline of the vertebrae. This set of images has a resolution of 0.17mm per pixel (Fig. 3c).

Examples of these three datasets and their manual landmarks are given in Fig. 3. In all experiments the manual annotations are only used to evaluate the performance of the method.

We compare the method with the two single-model initialisation strategies proposed in [6] and [7]. Given a large number of candidate parts, one approach constructs many different parts+geometry models of fixed number of parts, and uses a variant of Genetic Algorithm (GA) to select the best model [6], while the other builds a number of models of random configuration and adopts a voting strategy to achieve the same goal [7]. To generate the candidate parts, both approaches randomly select a reference image, place an overlapping grid on the reference and build a part model for each patch in the grid. In this way, we constructed over 1000, 2700 and 600 candidate parts for the fly wings, hands and spines, respectively. For each set of candidate parts, we first ran the voting based method to choose the optimal parts+geometry model and used it for initialisation. The resulting model determines the number of parts to be used in the GA based approach, from which we obtained a group of N_G models. The very best one was used to do the single-model initialisation as done in [6] and all of them were used for multi-model initialisation. In all experiments we set $N_G = 10$.

To evaluate the accuracy of the groupwise registration we compare with a manual annotation. We used the same protocol with [6] to compute the registration error. Point-to-point location error is reported for the fly wings and hands, and point-to-curve location error is reported for the spines. Results are given in Table 1. We used Welch's (two-tailed) t-test to compare the multi-model strategy with each single-model one, resulting in the p-values. The multi-model initialisation scheme performs similarly to the single-model ones on the fly wings, and does much better on the hands and spines. The power of multiple models can be clearly seen from the result on the spines, where the improvement is the most significant compared with what has been achieved on the other two datasets. This is not surprising as the spine dataset contains large shape variations and fractured vertebrae, which are very challenging for a single model to handle.

To further demonstrate the performance of the multi-model scheme, we repeated the above experiment with another 9 different reference images, thus totally 10 different

Table 1. Point location errors of the dense correspondence

(a) Fly wings (pixels)

Method	Parts	Mean±s.e.	Med.	90%	p-value
GA	20	1.9±0.06	1.7	2.7	$p < 0.01$
Voting	20	1.3±0.05	1.2	1.8	$p = 1$
Multi	20	1.3±0.04	1.2	1.7	–

(b) Hands (mm)

Method	Parts	Mean±s.e.	Med.	90%	p-value
GA	45	1.4±0.1	1.0	2.1	$p < 0.01$
Voting	45	1.3±0.09	1.0	2.2	$p < 0.05$
Multi	45	**1.1±0.04**	**1.0**	**1.8**	–

(c) Spines (mm)

Method	Parts	Mean±s.e.	Med.	90%	p-value
GA	21	3.9±0.5	2.6	4.7	$p < 0.01$
Voting	21	3.6±0.4	2.3	6.5	$p < 0.01$
Multi	21	**2.1±0.2**	**1.6**	**3.1**	–

Table 2. The influence of choice of reference images on the single-model and multi-model initialisation strategies

(a) Fly wings (pixels)

Method	$\hat{c}$	MAD	c_{min}	s.d.
GA	1.8	0.3	1.3	0.5
Voting	1.2	0.09	1.1	0.2
Multi	1.4	0.1	1.2	0.4

(b) Hands (mm)

Method	$\hat{c}$	MAD	c_{min}	s.d.
GA	0.9	0.07	0.8	0.5
Voting	0.9	0.06	0.8	0.6
Multi	0.9	0.04	0.9	0.3

(c) Spines (mm)

Method	$\hat{c}$	MAD	c_{min}	s.d.
GA	2.5	0.6	1.8	6.0
Voting	5.0	2.5	2.3	4.0
Multi	1.8	0.4	1.6	2.0

sets of experiments. This is to show the effects of different choice of the reference image. We computed the median $\hat{c}$ of the 10 medians c_i and the mean absolute difference MAD $= \sum_{i=1}^{10} |\hat{c} - c_i|/10$ for each method on each dataset. We compared $\hat{c}$ with the best median c_{min} of the 10 cases and MAD with the corresponding standard deviation. We summarise the results in Table 2. We find that the choice of reference images only has a small effect on the results of the multi-model initialisation strategy for all three datasets. Although a similar pattern can be observed for both single-model initialisation methods on the fly wings and hands, the performance of the two methods varies more on the spines. For example, the performance of the voting based method varies dramatically from one reference image to another.

4   Discussion and Conclusions

We have described a method that can effectively initialise groupwise non-rigid registration. This is achieved by using multiple parts+geometry models. Experiments show

that our algorithm is able to achieve state-of-the-art results, significantly outperforming earlier approaches that only use a single model. We also compared this multi-model initialisation strategy with the single-model one in terms of the influence of reference images. We find that the multi-model scheme is the least sensitive to the choice of reference images, suggesting that a robust system can be expected.

Current work indiscriminately uses the top models output by the GA based method [6]. If some models result in too many poor matches, the performance of the multi-model scheme will be inevitably degraded. Moreover, if different models share too many common parts, redundancy will arise. This will dilute the advantage of using multiple models to do initialisation, since different models may fail on the same images so that there is no chance to rectify those failures. In the future we will explore how to effectively choose a good set of parts+geometry models. The approach has a natural extension to 3D, which we intend to investigate.

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. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6), 681–685 (2001)
3. Frangi, A.F., Rueckert, D., Schnabel, J.A., Niessen, W.J.: Automatic 3D ASM Construction via Atlas-Based Landmarking and Volumetric Elastic Registration. In: Insana, M.F., Leahy, R.M. (eds.) IPMI 2001. LNCS, vol. 2082, pp. 78–91. Springer, Heidelberg (2001)
4. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23(suppl.1), 151–160 (2004)
5. Cootes, T.F., Twining, C.J., Petrović, V., Babalola, K.O., Taylor, C.J.: Computing accurate correspondences across groups of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(11), 1994–2005 (2010)
6. Zhang, P., Adeshina, S.A., Cootes, T.F.: Automatic Learning Sparse Correspondences for Initialising Groupwise Registration. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 635–642. Springer, Heidelberg (2010)
7. Zhang, P., Cootes, T.F.: Automatic Part Selection for Groupwise Registration. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 636–647. Springer, Heidelberg (2011)
8. 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)
9. Adeshina, S.A., Cootes, T.F.: Constructing part-based models for groupwise registration. In: Proceedings of International Symposium on Biomedical Imaging, pp. 1073–1076 (2010)
10. Fischler, M.A., Elschlager, R.A.: The representation and matching of pictorial structures. IEEE Transactions on Computer 22(1), 67–92 (1973)
11. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. International Journal of Computer Vision 61(1), 55–79 (2005)
12. Donner, R., Micusik, B., Langs, G., Bischof, H.: Sparse MRF appearance models for fast anatomical structure localisation. In: Bhalerao, A., Rajpoot, N. (eds.) Proceedings of British Machine Vision Conference, vol. 2, pp. 1080–1089 (2007)
13. 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)