Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2009 ; 5193140: 686–689. doi:10.1109/ISBI.2009.5193140.

Improving the Reliability of Shape Comparison by Perturbation

Y. Jiang¹, E. Edmiston², F. Wang², H. P. Blumberg^{1,2}, X. Papademetris^{1,3}, and L. H. Staib^{1,3,4}

¹Department of Diagnostic Radiology, Yale University, New Haven, CT, USA

²Department of Psychiatry, Yale University, New Haven, CT, USA

³Department of Biomedical Engineering, Yale University, New Haven, CT, USA

⁴Department of Electrical Engineering, Yale University, New Haven, CT, USA

Abstract

Shape comparison is a key scenario in morphometric study, where registration is often involved and found to be unreliable: different registrations can lead to different shape differences. This paper proposes a generic scheme applicable to most registration methods, to reduce this unreliability. It perturbs the registration processes by feeding them with resampled shape groups, and then aggregates the results to yield the final result. This scheme can be simplified for pair-wise registration methods to reduce the computation. Experiments are conducted on both synthetic and biomedical shapes using different registration methods, which demonstrate its effectiveness.

Index Terms

Shape registration; shape comparison; morphometric study; significance testing

1. Introduction

A key scenario in morphometric analysis is shape comparison, to compute the morphological difference between two groups of shapes. There are three basic modules involved in shape comparison: (1) shape registration; (2) the selection and computation of localized shape descriptors [1], *e.g.* landmark coordinates, deformation fields, or the Jacobian determinants of deformation fields; and (3) quantifying the shape differences, *e.g.* by significance testing [2] or classifier estimation [3]. Since it is difficult to find a practical classifier in many real applications, significance testing has been adopted widely [4], where the resulting *p* values are used to quantify the difference.

The shape registration module needs to do both shape registration and alignment. Registration finds landmarks which correspond across all the shapes, and alignment uses these landmarks to eliminate the similarity transformation between any two shapes. While alignment is a fairly mature process where different methods yield similar results [5] (Generalized Procrustes Analysis is adopted in this paper), registration currently shows much more unreliability where different methods often lead to different correspondences and consequentially shape differences. This may not be surprising because as an ill-posed problem, its solution really depends on the prior or regularization that is assumed in each method. How to reduce this variability is a serious question since shape comparison has started to be used in clinical research, but to our best knowledge, this question has seldom been discussed so far.

In this work we develop a generic scheme which can be applied to different registration methods to improve their reliability. The intuition is that most registration processes are largely

influenced by either their parameters or the data being registered. Such unstable processes have a good chance of improving by perturbing them and aggregating the results, which has already been a successful strategy adopted in machine learning [6]. In this paper, Section 2 introduces the proposed perturbation scheme, goes over a conceptual justification, and describes a detailed implementation; Section 3 demonstrates some preliminary experimental results on both synthetic and real biomedical data; and Section 4 concludes the paper.

2. Method

To perturb a registration process, we can either perturb its algorithm parameters (note that initial conditions can also be considered as algorithm parameters) or resample the data. Resampling data also perturbs the registration since most registration algorithms are data-driven. In this paper, we choose the later because it is difficult to design a unified scheme to perturb parameters for different algorithms, and deviating parameters from their optima for any given data is unlikely able to make positive contributions in aggregation.

The results of perturbation will be evaluated in the scenario of shape comparison since improving the reliability of shape comparison is our ultimate goal. In this paper, we follow the simplest way for shape comparison: taking the coordinates of registered landmarks as descriptors, and applying Hotelling's T^2 -testing for each set of corresponding landmarks to find a p value as the measure of local difference. The evaluation is very straightforward when the ground truth of correspondence are given (thus the shape difference). In this paper, a group of bump boxes are synthesized for this purpose. For real biomedical data whose true difference is never known, we compare the results from different registration methods. It is considered that the more similar their results are, the more reliable the computation is – similar criteria have been adopted to evaluate image registrations.

2.1. Why perturbation works

Given a group of shapes $\mathcal{S} = \{\mathbf{s}_n, n = 1, \dots, N\}$, the goal of shape registration is to find the correspondence across all $\{\mathbf{s}_n\}$. One explicit way to represent the correspondence is locating a set of corresponding landmarks $\mathbf{x}_n = \{\mathbf{x}_{n,m} = (x_{n,m}, y_{n,m}), m = 1, \dots, M\}$ for each shape $\mathbf{s}_n$. For simplicity, we assume there exists a mean shape $\mathbf{s}_0$ whose landmarks are given and fixed as $\mathbf{x}_0$. In this case, the registration process is for each $\mathbf{s}_n$ to find M landmarks $\mathbf{x}_n$ which correspond to $\mathbf{x}_0$.

Most group-wise registration methods are data-driven. It is also true for pair-wise registration methods because the template shape needs to be selected from the given group, normally by choosing the one that has least deformation to all the others. If we denote the correspondence resulting from a certain registration process as ϕ , we can write

$$\mathbf{x}_n = \phi(\mathbf{s}_n, \mathcal{S}). \quad (1)$$

The perturbation adopted in this paper can be considered as acquiring a sequence of shape groups $\{\mathcal{S}_k, k = 1 \dots K\}$ each consisting of N independent observations from the same underlying distributions as $\mathcal{S}$. In practice, we only have a single shape group $\mathcal{S}$ thus $\{\mathcal{S}_k\}$ are normally imitated by bootstrap samples $\{\mathcal{S}_k^{(B)}\}$. After registration for all these groups, the simplest way for aggregation is to replace $\phi(\mathbf{s}_n, \mathcal{S})$ by the average of $\phi(\mathbf{s}_n, \mathcal{S}_k)$ over k , which ideally is $\phi_A(\mathbf{s}) = E_{\mathcal{S}}\phi(\mathbf{s}, \mathcal{S})$, where $E_{\mathcal{S}}$ denotes the expectation over $\mathcal{S}$.

The advantage of $\phi_A(\mathbf{s})$ over $\phi(\mathbf{s}, \mathcal{S})$ is simply based on a general inequality, which is $EX^2 \geq (EX)^2$. Assume the joint distribution of $(\mathbf{x}, \mathcal{S})$ to be P once $\mathbf{x}_0$ is given. Let $(\mathbf{x}, \mathcal{S})$ be independently

drawn from P , and take $\mathbf{s}$ to be a fixed input value and $\mathbf{x}$ an output value. Note that $E_S \phi(\mathbf{s}, \mathcal{S}) = \phi_A(\mathbf{s})$, by the above general inequality we have

$$E_S (\mathbf{x} - \phi(\mathbf{s}, \mathcal{S}))^2 \geq (\mathbf{x} - \phi_A(\mathbf{s}))^2. \quad (2)$$

Integrating both sides of (2) over the joint distribution of $(\mathbf{x}, \mathbf{s})$, we get that the mean-squared error of $\phi_A(\mathbf{s})$ is lower than the mean-square error averaged over $\mathcal{S}$ of $\phi(\mathbf{s}, \mathcal{S})$, which means statistically $\phi_A(\mathbf{s})$ is better than $\phi(\mathbf{s}, \mathcal{S})$. How much better depends on the degree of inequality, which is apparently affected by the instability of $\phi(\mathbf{s}, \mathcal{S})$. The more variable the $\phi(\mathbf{s}, \mathcal{S})$ are, the more aggregation may improve.

Once the correspondence is established, the inter-group statistical shape difference, which is defined as the p value $\{p_m, m = 1 \dots M\}$ for landmarks $\{\mathbf{x}_{n,m}\}$, is also determined with a significance testing routine, which can be written as $\{p_m\} = \mathbf{P}(\{\mathbf{x}_{n,m}\}) = \mathbf{P}(\{\phi(\mathbf{s}_n, \mathcal{S}), \mathbf{s}_n \in \mathcal{S}\})$. By applying the same inequality as above, we also have

$$\begin{aligned} & E_S [\mathbf{P}(\{\mathbf{x}_n\}) - \mathbf{P}(\{\phi(\mathbf{s}_n, \mathcal{S}), \mathbf{s}_n \in \mathcal{S}\})]^2 \\ & \geq [\mathbf{P}(\{\mathbf{x}_n\}) - E_S (\mathbf{P}(\{\phi(\mathbf{s}_n, \mathcal{S}), \mathbf{s}_n \in \mathcal{S}\}))]^2, \end{aligned} \quad (3)$$

which suggests that aggregating the correspondences and aggregating the testing results both have a chance to improve the computation of shape difference. The latter is adopted in this paper due to its simplicity.

2.2. Implementation

Selecting the mean shape $\mathbf{s}_0$ —We chose a typical shape $\mathbf{s}'_0$ to approximate $\mathbf{s}_0$, and simply take all its points to be $\mathbf{x}_0$. This shape is chosen to have the least deformation to all the others, in the same way as selecting the template shape for pair-wise registration algorithms. Since we need to establish the correspondence between any $\mathbf{s}_n$ and $\mathbf{s}_0$, $\mathbf{s}'_0$ will join all the resampled groups $\{\mathcal{S}_k^{(B)}\}$ to link different registrations. Furthermore, because the resulting landmarks on $\mathbf{s}'_0$ after registering $\mathcal{S}_k^{(B)}$, say $\mathbf{x}_0^{reg}$, would never automatically locate exactly at $\mathbf{x}_0$, we need to get the correspondence on $\mathbf{x}_0$ by interpolation. As mentioned at the end of Section 2.1, we actually interpolate the significance testing results on $\mathbf{x}_0^{reg}$ onto $\mathbf{x}_0$.

Resampling data for $\{\mathcal{S}_k^{(B)}\}$ —Suppose we are computing the shape difference for two groups $\mathcal{S}^a$ and $\mathcal{S}^b$, where $\mathcal{S} = \{\mathcal{S}^a, \mathcal{S}^b\}$. Each $\mathcal{S}_k^{(B)}$ will be formed by N^a samples from $\mathcal{S}^a$ and N^b samples from $\mathcal{S}^b$, both of which are obtained by bootstrapping, *i.e.* randomly sample the original group with replacement. This process is repeated K times.

Aggregating perturbation results—After registering all $\{\mathcal{S}_k^{(B)}\}$, we now have K sets of p values for each landmark $\mathbf{x}_{0,m}$, $\{p_{m,k}, m = 1, \dots, M, k = 1, \dots, K\}$. In this paper, we aggregate using the following methods:

1. Classify each $\{p_{m,k}\}$ to be significant or not significant for a given significance level α , and assign the final classification by a majority vote.

2. Simply let $p_m = \frac{1}{K} \sum_{k=1}^K p_{m,k}$, as described in Section 2.1. We denote it as p_m^{A1} . Since p is highly nonlinear, we also tested $p_m^{A2} = \text{median}(\{p_{m,k}\})$.

In the first case, we do not have a p value for each landmark; instead a label l_m is obtained, which still enables the comparisons between shape differences.

Comparing the resulting shape differences—To compare various shape difference $\{p_m, m = 1, \dots, M\}$ we count the number of p_m s that are both significant or insignificant in different results, according to a given significance level α , *i.e.*,

$$\begin{aligned} \text{sim}(\alpha) &= \frac{1}{M} (nr(\{m, p_m^{reg1} > \alpha\} \cap \{m, p_m^{reg2} > \alpha\}) \\ &\quad + nr(\{m, p_m^{reg1} \leq \alpha\} \cap \{m, p_m^{reg2} \leq \alpha\})), \end{aligned} \quad (4)$$

where the superscripts of p indicate the registration approach. In practice, α is typically 0.05. In this paper, we compute the similarity over $\alpha \in (0.01, 0.1)$. For the first aggregation method above, we simply count the number of common labels between shape differences.

2.3. Simplification for pair-wise registration methods

Shape registrations are time consuming since they are often difficult optimization problems. The proposed scheme requires registration to be carried out for the resampled groups for K times, $K > 20$ in this paper. However, time can be saved for pair-wise registration methods, where the proposed scheme is approximated by only perturbing the registration templates. This is because pair-wise registrations are only affected by template shape, which simplifies Eq.(1)

to be $\mathbf{x}_n = \varphi(\mathbf{s}_n, \mathbf{s}_t^S)$, where $\mathbf{s}_t^S$ is the template shape selected from $\mathcal{S}$. Assuming

$\mathbf{P}(\{\varphi(\mathbf{s}_n, \mathbf{s}_t^S), \mathbf{s}_n \in \mathcal{S}_k\}) \approx \mathbf{P}(\{\varphi(\mathbf{s}_n, \mathbf{s}_t^{S_k}), \mathbf{s}_n \in \mathcal{S}\})$, which is reasonable when $\mathcal{S}_k$, as resampled by bootstrapping from $\mathcal{S}$, has a similar distribution to $\mathcal{S}$, we have

$E_s[\mathbf{P}(\{\varphi(\mathbf{s}_n, \mathbf{s}_t^{S_k}), \mathbf{s}_n \in \mathcal{S}_k\})] = \sum_{t=1}^N \frac{w_t}{\sum_{t=1}^N w_t} \mathbf{P}(\{\varphi(\mathbf{s}_n, \mathbf{s}_t), \mathbf{s}_n \in \mathcal{S}\})$, where w_t is the frequency that $\{\mathbf{s}_t, t = 1 \dots N\}$ is selected as the template shape in all $\{\mathcal{S}_k, k = 1 \dots K\}$. As a matter of fact, $w_t \approx 0$ for most $t = 1 \dots N$, which makes registration required (for the original shape group with different template) much less than K times.

3. Experimental Results

3.1. Registration methods

For 2D shapes, we tested the proposed scheme on three typical registration methods [7–9], which assume different priors on the shape transformation. The prior in [7](MDL) is based on information theory, while in [8](SCM) and [9](IMP) certain splines are used, either global (thin plate splines) or local (B-splines). Both MDL and SCM are group-wise registration methods, while IMP is a pair-wise one. For 3D shapes, we converted the shape data into images by *distance transform* and registered them by three pair-wise algorithms implemented in Bioimage Suite (www.bioimagesuite.org), which adopt the normalized mutual information (NMI), the sum of squared difference (SD), and correlation ratio (CR) as similarity measures respectively, but all use free form deformation (FFD) for transformation.

3.2. Results on 2D synthetic shapes

The synthetic shapes consist of two groups of bump boxes as shown in Fig 1, whose true correspondence is given. Some results are shown in Fig. 2, where $N^a = N^b = 6$ and $K = 24$ for

perturbation. It is observed that for all registration methods, the perturbation scheme has resulted in shape differences that are obviously closer to the ground truth with the exception of the majority voting and p^{A2} for MDL. This exception probably implies that this method is less stable, and only a minor part of its results are close to the ground truth. The magnitude of improvement on p^{A1} also indicates that MDL benefits most from the perturbation among all the three registration methods, which again implies its instability. For pair-wise registration method IMP, perturbing the registration template achieved similar improvement compared with the proposed scheme (cf (h)(i) in Fig. 2), as expected in Section 2.3.

It is also observed (not shown here) during experiments that the correspondences established by shape registration all look good by visual inspection. The shape generalization analysis results are also found to be very similar to that obtained by the ground truth. However, large areas of incorrect shape difference are observed, especially for the results of MDL and SCM. This suggests that the inter-group shape difference could set a stricter or more sensitive criterion for shape registration, compared with the criteria in generalization analysis. One reason is that the difference here is defined by a p value from significance testing, which is not decided by the magnitude of any error, but only by statistical significance.

3.3. Results on 3D biomedical shapes

We computed shape differences between two groups of 3D brain structures, each containing 6 pairs of hippocampi. We already know that the first group has larger overall volumes, but we want to know which parts of the structure have significant local difference. The data are obtained from [10].

We compute the similarities among the resulting shape differences from the three registration methods described in Section 3.1. As shown in Fig. 3, large differences are observed between the results from different registration methods. But this difference is very much reduced after the perturbation scheme is employed. This indicates that the perturbation scheme leads to a more reliable computation of shape difference. Similar results are observed on a set of 2D femur profiles and a group of 2D corpus callosum shapes.

4. Conclusion

A perturbation scheme has been shown to be able to improve the reliability of shape comparison. It can be applied to any existing registration method by simply adding a loop in front to feed different resampled groups and a back end that aggregates all the results. It works for both pair-wise and group-wise methods, and for pair-wise methods, it can be further simplified to reduce the computation.

Acknowledgments

Thanks to the anonymous reviewers for their comments. This work was supported by the NIH under grants R01EB006494 (XP), R01MH070902 (HPB), R01MH069747 (HPB), and R01NS35193 (LHS, XP).

References

1. Adams DC, Rohlf FJ, Slice DE. Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology* 2004;71(1):5–16.
2. Ashburner J, Friston KJ. Voxel-Based Morphometry-The Methods. *Neuroimage* 2000;11(6):805–821. [PubMed: 10860804]
3. Golland P, Fischl B. Permutation tests for classification: Towards statistical significance in image-based studies. *IPMI* 2003:330–341.

4. Zhou L, et al. A study of hippocampal shape difference between genders by efficient hypothesis test and discriminative deformation. MICCAI (1) 2007:375–383.
5. Timoner SJ, et al. Performance issues in shape classification. MICCAI (1) 2002:355–362.
6. Breiman L. Bagging Predictors. Machine Learning 1996;24(2):123–140.
7. Thodberg, HH. Proc IPMI. BMVA; 2003. Minimum description length shape and appearance models; p. 51–62.
8. Chui H, Zhang J, Rangarajan A. Unsupervised learning of an atlas from unlabeled point-sets. IEEE Trans Patt Anal Mach Intell 2004;26:160–173.
9. Huang X, Paragios N, Metaxas DN. Shape registration in implicit spaces using information theory and free form deformations. IEEE Trans Patt Anal Mach Intell 2006;28(8):1303–1318.
10. Yang J, Staib LH, Duncan JS. Neighbor-constrained segmentation with level set based 3-d deformable models. IEEE Trans Med Imaging 2004;23(8):940–948. [PubMed: 15338728]

Fig. 1.

Each bump box has three semi-circles on the different edges of a rectangle box (some noise also imposed). The bump locations, $\{P1, P2, P3\}$, are different from sample to sample and uniformly distributed along the edges. Radius $R3$ is identical for both groups, while $R1$ and $R2$ are different. Each group consists of 9 different shapes. Right figure shows typical samples from both groups.

Fig. 2.

Bump box results. (a)-(e) Ground truth and results obtained by IMP [9]. (a)(b) Correspondence between two typical shapes. (c) shows the true shape differences mapped to S'_0 , (d) denotes the shape difference by registering once and (e) by perturbation. (f)-(h) Similarity between truth shape differences and results by registrations, with or without perturbation.

Fig. 3.

Real biomedical data results. (a) Three shape differences of 3D hippocampus, resulting from three similarity metrics during registration: NMI, SSD, and CR. from left to right. The first row shows the results by registering once while the second row shows the result by perturbation, $\{p^{A1}\}$. (b) Similarity of the shape differences resulting from these three similarity metrics, as a function of significance level.