Relative Statistical Performance of S-reps with Principal Nested Spheres vs. PDMs

Stephen M. Pizer, Junpyo Hong, Sungkyu Jung, J.S. Marron, Jörn Schulz**, Jared Vicory*
Univ. of North Carolina, *Univ. of Pittsburgh, **Tromsø Univ.

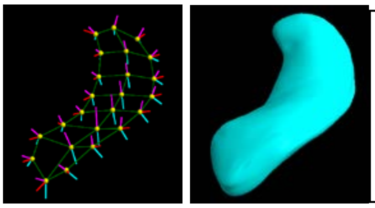

Fig. A mean s-rep for a hippocampus and its implied boundary. Balls = skeletal points. Spokes proceed from the balls.

Aims: The single-figure discrete quasi-medial skeletal representation of anatomic objects called the *s-rep* (Fig.), capturing position (via skeletal samples), orientation (of the quasi-boundary-normals forming the s-rep spokes), and width properties (spoke lengths), has been used for a variety of goals of or using shape statistics. We summarize the performance of these statistics and compare them to those on the more prevalent object representation of boundary point distribution models (PDMs).

Methods: Methods for probability distribution estimation, classification, and hypothesis testing have been developed using s-reps. All involve recognizing that for both s-reps and PDMs, many of their geometric properties abstractly live on spheres and thus, when analyzing them statistically, the method of Principal Nested Spheres (PNS) [Jung] is needed to statistically preprocess these properties before applying methods designed to work in Euclidean spaces.

1) Estimating a probability distribution on objects, incl. modes of variation, principal variances, and total variance, using (Euclidean) PCA. We show how to transform incommensurate shape variables into comparable units, needed for PCA and allowing comparison of methods as to total variance. Also, we show the necessity of fitting s-reps to primary object descriptions in correspondence, which requires estimating the probability distribution. In our method the correspondence was produced by a final stage of s-rep fitting that initialized from a common mean and was fitted over common modes of variation, followed by spoke length refinement.
2) Training and execution of classification between 2 classes of objects using (Euclidean) DWD [Marron], less sensitive to noise than SVM. Histograms in DWD's separation direction in the Euclideanized feature space were used by a Bayesian approach to produce the function P(schizo | position along the separation direction). Varying the prior probability of being schizophrenic yielded a curve: true positive rate vs. true negative rate.
3) Hypothesis testing to discriminate geometric properties that differ significantly between two classes of objects, using (Euclidean) permutation tests on geometric properties with locality, via family-wise error rate to correct for multiple tests. The permutation subclass means were computed from the backward means implied by PNS, and the inter-mean distances for the subclasses were computed geodesically for sphere-resident geometric properties.

Results: Our data consists of hippocampi segmented from MRIs. 221 are from first-episode schizophrenics, and 56 are from controls. The PDMs were constructed as the boundary points at the end of the s-reps spokes. We made the following comparisons:

1) Classification into schizophrenic and normal classes, comparing via Area under ROC, using iterated cross-validation [Hong]. Results: s-reps with Euclideanization: 0.65, s-reps without Euclideanization: 0.56, PDMs with Euclideanization: 0.63, volume only (the common neuroscience approach): 0.58. Visualization of the s-reps falling along the separation vector through the population mean of the pooled classes yields comprehension of the inter-class shape differences.
2) Estimating a probability distribution [Hong paper in preparation]. First, we found s-reps to have 9% more total variance per dimension than the PDMs. Second, PDMs required 10 eigenmodes to obtain an accumulated variance per dimension equal to 80% of the s-reps' total variance per dimension, whereas the s-reps required only 4 eigenmodes. When doing the statistics without correspondence, 25 eigenmodes were required.
3) Hypothesis testing on the s-reps [Schulz] showed significant global differences between the classes. It also showed significant differences in skeletal position and s-rep spoke orientation for specified spokes.

Conclusion: With respect to this data, statistical analysis via Euclideanization of s-reps is superior.

References: Hong, JP, J Vicory, J Schulz, S Jung, M Styner, JS Marron, S Pizer Classification of medically imaged objects via s-rep statistics. http://midag.cs.unc.edu/MIDAG_FS.html, submitted.
Jung, S, I Dryden, JS Marron, ``Analysis of Principal Nested Spheres" , *Biometrika*, **99**(3), 551-568 (2012).
Marron JS, M Todd, J Ahn. Distance-weighted discrimination. *J. Am. Stat. Assoc.* **102**(480): 1267-1271 (2007).
Schulz, J. Stat'l anal. of med. shapes and directional data. Dissertation, Arctic Univ. of Norway, 2013.

Acknowledgements: Others, too numerous to name, while at UNC have contributed to this research over many years. The Lilly Foundation provided the base hippocampus MRIs. Martin Styner provided the segmented hippocampi and useful advice.