Templates: a complex implicit parameter for morphometric analyses takes the form of an algorithmic flow

Fred L. Bookstein

University of Vienna, University of Washington

The multivariate statistics of *landmark points,* configurations of discrete points with individual characterizations, has been well-established now for over a decade. Versions of this approach via the tangent space to Kendall shape space exist for group comparisons, principal components, predictions of effects, and trends over exogenous causes, along with special techniques altering the basis of the linearized space for specific biological questions such as symmetry, which has been the topic of papers at recent LASR meetings, or allometry (the dependence of shape on size).

But landmark configurations do not make up the full resource of data for shape statistics in its current biomedical application domains. For 3D data, they are not even the major component of those data resources. With the help of the thin-plate spline, European morphometrics (or at least anthropometrics) has now turned to the more congenial formalism of *semilandmarks,* “sliding landmarks” that represent information from identifiable curves on the surface and from regions of the surface in-between the landmarks or curves (Gunz et al., 2004). Landmarks can emerge as subordinate features, such as vertex points, of the 2D or 3D curves that serve as the primary data, and 3D curves, in turn, can emerge as descriptive summaries of extended surfaces, such as the *ridge lines* that integrate vertices of normal sections along a principal curvature.

Semilandmarks, once computed in a principled way, can be treated as proper landmarks through all of the usual Procrustes shape space formalisms. This strategy relies on an assumption of coherent preprocessing — careful management of all the dimensions that were adjusted at the time of the original digitizing. Their variation is not precisely zero, as they depend on derivatives of the sample of actual forms, but it is enormously attenuated in comparison to variations compatible with the mean constraints. Communications to the scientific community about findings based on semilandmarks surely go best in the original figure space, with all coordinates, constrained or not, shown in place. There is thus a need for formal protocols that appropriately constrain the linearized (tangent-space) configurations for Procrustes operations on the redundant representation.

Earlier statistical approaches to these themes are sparse. One version (Katina and Bookstein, 2008) uses landmarks to drive an unwarping to the Procrustes average shape, then treats the remaining variability at semilandmarks as if the generalized least-squares procedure that generated it was orthogonal to the Procrustes metric rather than skew. A formal inferential approach (Mardia, Kirkbridge, and Bookstein, 2004) limits its attention to the case of one single landmark with a tangent line to a curve through it, the “edgel.” A formalism is needed instead that will pass directly from the simplicial complex of the landmark-curve-surface scheme to the appropriate Procrustes tangent space. This note suggests that that formalism be algorithmic, not algebraic: that the enunciation of point-curve-surface semilandmark schemes take the form of orchestrated sequences of a modest variety of image matching/warping operations. These atomic operations are construction of privileged planes (symmetry plane, nongeneric tangent planes), intersections with planes, location of points, location of ridge curves, warping by partial information, and pointwise projection post-warp orthogonally onto curves or surfaces. The overall bending-energy minimization explicit in the theorems underlying the semilandmark algebra of the single spline is now inoperative, but for samples of more than two forms it was inoperative anyway, owing to the nonlinearity with which variation of the specimen surface normals enters into the bending energy.

I illustrate these considerations by an example that I hope will prove typical: a template for a configuration of 15 landmarks, 127 curve semilandmarks, and 273 surface semilandmarks from a study of the surface of the growing human mandible as seen in CT scans of dry dead specimens. The example appears by the kindness of Mr. (soon to be Dr.) Michael Coquerelle of the University of Toulouse, working under the supervision of José Braga (Toulouse), Gerhard Weber and myself (Vienna), and Demetrios Halazonetis (Athens). The specimens under study range in age from prenatal to about six years. Their number of erupted teeth is thus quite variable.

Otherwise explicit formulas become implicit via a "digitizing flowchart," a sequential set of instructions — a constructive morphometric geometry — according to which a template is very carefully constructed on a single a-priori "typical" specimen. The same constructions control the task of "digitizing" every additional specimen. In this setting, the result of a digitization is a carefully controlled warp of the entire template into the space of each target specimen in turn. The figure page shows a variety of views of this one single template. In the slides as projected, and in any version of this note that I might send you in response to requests, the surface extracted on the original CT scan is colored red. The green crosses with labels are conventional landmark points. The white polylines are curves, and the green crosses without labels are curve semilandmarks or surface semilandmarks. Landmarks are or are not bound to curves, which in turn may or may not be bound to surfaces, according to incidence properties to be reviewed below.

The constructive geometry begins with two specific planes:

P1. A subjectively located "midsagittal plane" around which the surface is most nearly symmetric as viewed by eye. For instance, the plane passes through the saddles in the alveolar ridge between the two lower front incisors. It is *not* the mirroring plane generated by Procrustes superposition of the form upon its own mirror image, because the average normal mandible is not quite bilaterally symmetric.

P2. A best-fit plane tangent to the upper margin of the mandible. This plane is typically in approximate contact with the mandibular surface at four separate points CorL, CorR, and the two condylar tops ConT as shown. These structures typically lie very close to a plane owing to the symmetry required for efficient biomechanics of chewing. They comprise our first four landmarks.

From the interaction of P1 with the surface we extract one particular curve:

C1. The *symphysis* is the plane curve cut by the midsagittal plane P1 upon the surface.

This curve includes two further landmark points extracted by hand once the intersection of the plane P1 with the mandibular surface is visible: Inf1, Inf2, lingual and buccal vertices of the curve C1 near the incisors.

It is convenient at this time to augment the list of landmarks by four additional points MenL, MenR, MandL, MandR, the paired mandibular foramina, which (see the figures) float in space over the apertures of those foramina where the corresponding tangent plane is actually "missing." Also produced are the points ConRL, ConRM, ConLM, ConLL where the surface normal of the condyle is perpendicular to the symmetry plane P1.

Following the general constructive theory of ridge curves (cf. Koenderink, 1990: preimages of the cuspidal edges of the surface of centers for the given surface) we extract three curves of this type by inspection of the corresponding normal sections.

C2. The *mandibular border* is the ridge curve along the bottom of the mandibular bone, from the midsagittal outward in both directions as far as it can be extended. On the *template,* production of this polyline is by careful handwork spacing forward along the tangent by distances inversely proportional to curvature and inspecting normal sections to ensure that the candidate point is a vertex of every curve of normal section.

Because the bilateral asymmetry of the mandible is one frequent topic of investigation, the morphometric geometry requires symmetric rosters of semilandmarks. In practice, one places them sensibly on one half of the form, then warps the form onto itself via reflected relabelling of landmarks and finally projects from the warp to the nearest point of the corresponding (sub)manifold on the other half. In effect, the "real" template consists only of one hemimandible (including the unpaired curve, C1); the other half is digitized as warp of the first, hence at constrained spacing.

C3, C4. Typically each coronoid point is quite nearly upon the *anterior ramus ridge curve* that runs from just behind the last molar to just in front of the condyle. We extract these curves, left and right. Again, for the template, the second of these curves is created by pointwise projection of the reflected relabelled warp of the first, the only one that could be digitized at free spacing.

The remaining landmark point Symph is produced as the intersection of the curves C1 and C2, a plane curve and a ridge curve. It is close to the point that orthodontists often name "Menton."

There remain two more curves, the *alveolar curves* C5, C6, which run along the rim of the alveolus where the erupting teeth cut it (always near the ridges of the alveolar surface itself, structures that are visible only where the teeth are absent). These are thus not ridge curves, but ordinary polylines made up of variable numbers of discretely sampled touching points. They will eventually be resampled like any other set of curve-bound semilandmarks.

Finally, the surface points of one side of the template, spaced pleasantly in a quasi-grid arrangement, are paired to their antimeres by projection downward from the warp driven by the reflected relabelling of all the landmark points and curves.

Thus the template. The application of this scheme to any other form proceeds in the same conceptual order interweaving planes, points, curves, more points, more curves, and surfaces. To digitize is to interweave point specification, warping, and orthogonal projection in arbitrarily complicated but fully specified sequences. Data digitized at any step constrain the forward warp of the template that applies to all subsequent steps. The digitizing is not order-independent, and so invariance of the scripted order is mandatory. At any stage, ridge curves are warped in from the template and then projected "down" to the visible ridges of the target surface (this is a simple operation given access to arbitrary surface sections, in this case, normal sections along the tangent to the ridge curve in question). Ridge curves can also be digitized by subdivided updating of the warp, beginning with the midpoints, then the quarter-points, and so on. Finally, after the surface semilandmarks are warped into range (so that projection variability is attenuated) by a spline driven by landmark points and curves, they too are projected all at once, down onto the target surface. The normal components of these projections are evaluated independently, so we are no longer precisely optimizing any global figure such as bending energy.

In this way the constraints underlying any particular semilandmark scheme remain implicit in the digitizing sequence. All of our advance tacit knowledge — the symmetry of the form, the four-point plane of upper paired structures, the reliable ridge curves, the special treatment of the alveolar ridges — must be explicitly coded in two sets of instructions, one for the template and one for its warping onto the target. Symmetry is ubiquitous. It is handled by explicit construction of a midsagittal (but not a mirroring) plane, by explicit indexing of paired vs. unpaired landmarks, and by explicit symmetrizing of semilandmarks that are themselves paired. The forward warp is constructed over and over: first from the landmarks that are easiest to locate without the warped template, then from the foramina and the semilandmarks on the condyles, then from the last landmark on C1, and finally from all the other ridge curves. The constructions keep carefully separate the incidence relations of points, curves, and surfaces. Ridge curves, of course, are necessarily on the surface. The alveolar curves are not ridge curves, but polylines.

Work by the EVAN Toolkit group (Paul O'Higgins, Roger Phillips, Oualid Ben Ali, William Green, and this author) will implement exactly this scripted data flow in a modification of `Edgewarp.` The new tool should be available for user experiments and extensions by late this year.

Here as in so many other applications of biomathematics to biometrics we see how much more information is coded in biological data sets than just in the coordinates inhabiting our data matrices. I put this example before you as one guide to a standard of care for morphometric archives that is comparable in its tuning to the standards for genomics and proteomics that are likely more familiar to readers of this Proceedings. When morphometrics is attached to biological theory it inherits implicit constraints as rich as does bioinformatics, and so it, too, will benefit from many of the same interdisciplinary attempts at jointly informed tools.

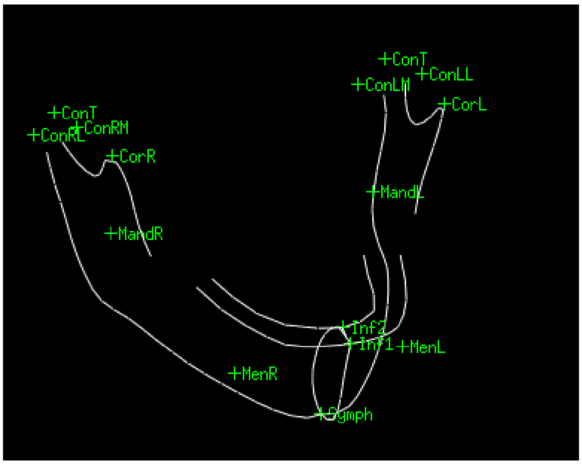
References

Gunz, P., Mitteroecker, P., and Boostein, F.L. (2004), Semilandmarks in three dimensions. In Slice, D. (ed.), *Modern Morphometrics in Physical Anthropology*, Springer

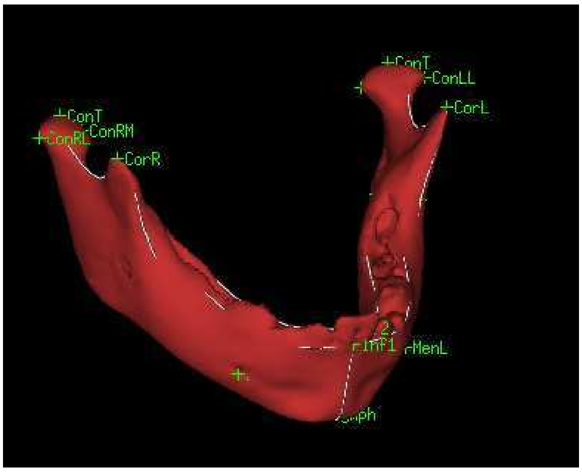
Katina, S. *et al.* (2008), Was it worth digitizing all those curves, *Manuscript in preparation*

Koenderink, J. (1990). *Solid Shape*, MIT Press

Mardia, K.V., Kirkbride, J., and Bookstein, F.L. (2004).Statistics of shape, direction and cylindrical variables, *Journal of Applied Statistics*, **31**:465–479

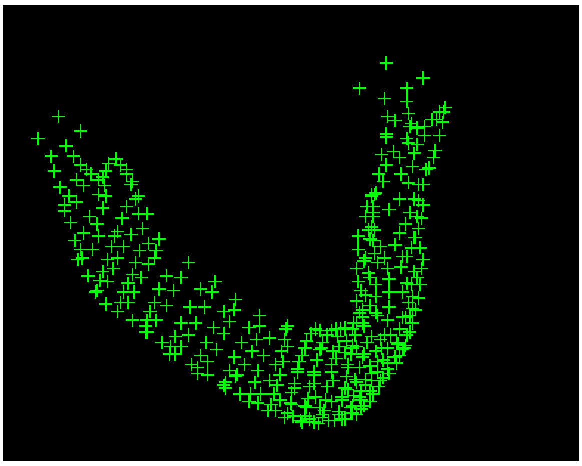


Landmark points and curves

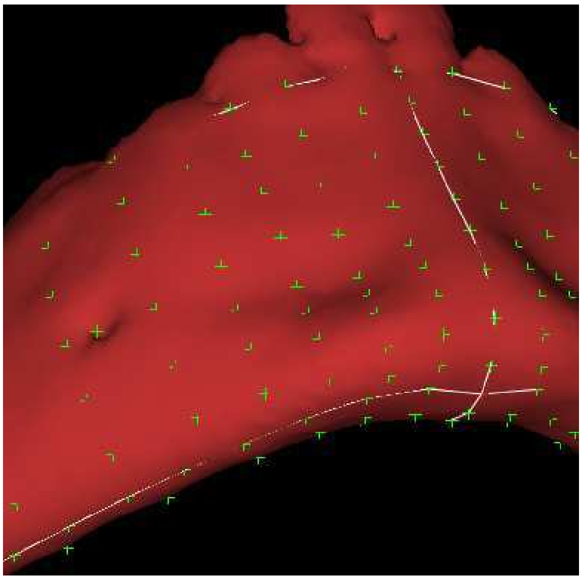


Landmarks, curves, surface

↑all points, curves, surface (detail↓)



the points alone (the Procrustes data resource)



Acknowledgements. I thank Philipp Gunz and Simon Neubauer, Max Planck Institute, Leipzig, and Gerhard Weber, Katrin Schäfer, Stanislav Katina and everybody else in the morphometrics group at the Department of Anthropology, University of Vienna, for exploratory conversations. An original version of the mandible template here was modified (to be symmetric) by Michael Coquerelle, who also provided the exemplary surface form used in the figures. We all were supported in part by the sixth European Union Framework Programme under Marie Curie contract 2005–01–109564 administered by the University of Vienna, and by Horst Seidler, chair of the anthropology department there. Figures were produced in `Edgewarp,` Dr. William Green's program package for spline-driven 3D medical image digitizing and navigating. A shorter version of this talk was presented at the American Association of Physical Anthropologists, Columbus, Ohio, April 12, 2008.