A Brief Introduction to Statistical Shape Analysis

Mikkel B. Stegmann and David Delgado Gomez*
Informatics and Mathematical Modelling, Technical University of Denmark
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark

6th March 2002

Abstract

This note aims at giving a brief introduction to the field of statistical shape analysis looked at from an image analysis point of view. Basic techniques such as Procrustes analysis, tangent space projection and Principal Component Analysis (PCA) are presented and subsequently demonstrated and discussed in depth through a biometric case study of hands.

Prerequisites are limited to basic knowledge of linear algebra and statistics at the price of a somewhat less compact presentation. To encourage learning by exploration; images, annotations and data reports from the hand study are made available for download.

Keywords: statistical shape analysis, landmark methods, morphometrics, Procrustes analysis, principal component analysis, multivariate statistics, biometrics, active shape models.

Contents

*Contact information: {mbs,ddg}@imm.dtu.dk, http://www.imm.dtu.dk/~mbs/

1 Introduction

For exactly a decade ago the image analysis community was introduced one of the more influential ideas within the community these past ten years. The paper giving name to this idea borrowed attention from the much famous *Snakes* paper [15] by Kass, Witkin and Terzopoulos and was published in 1992 as *Active Shape Models – 'Smart Snakes'* [4] and the later major introduction as *Active Shape Models - their training and application* [6] by T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham in 1995. The brotherhood with Snakes was rightfully claimed as they both are deformable models but contrary to Snakes, Active Shape Models (ASMs) have global constraints w.r.t. shape. These constraints are learned though observation giving the model flexibility, robustness and specificity as the model only can synthesize plausible instances w.r.t. the observations.

This note introduces the foundation of Active Shape Models[1], namely the statistical analysis of shapes – a discipline that dates back to the work of Sir D'Arcy Wentworth Thompson [19] – which is widely applicable to many other areas than image analysis. For further readings we have a few pointers into the vast amount of literature [7, 11, 1, 10, 5, 17].

2 Shapes and Landmarks

What do we actually understand by the concept of *shape*? In this text we will adapt the definition by D.G. Kendall [7]:

> **Definition 1**: **Shape** *is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object.*

According to this shape is – in other words – invariant to Euclidean similarity transformations. This is reflected in fig. 1. The next question that naturally arises is: How should one describe a shape? In everyday conversation, unknown shapes are often described as references to known shapes – e.g. *"Italy has the shape of a boot"*. Such descriptions can obviously not easily be utilized in an algorithmic framework.

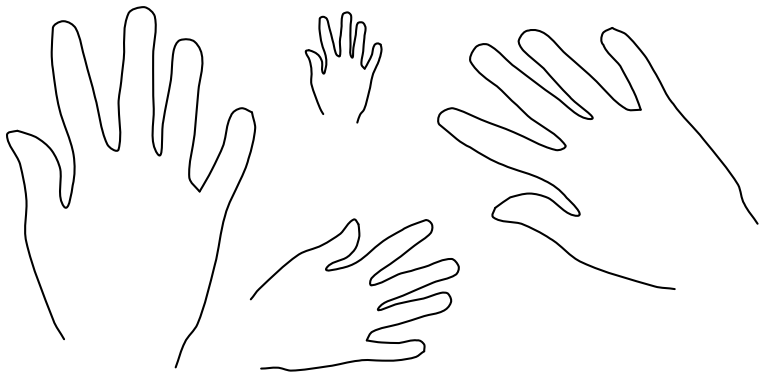


Figure 1: Four copies of the same shape, but under different Euclidean transformations.

One way to describe a shape is by locating a finite number of points on the outline. Consequently, the concept of a *landmark* is adapted [7]:

> **Definition 2**: *A* **landmark** *is a point of correspondence on each object that matches between and within populations.*

[1] Beside ASMs statistical shape analysis is also the foundation the myriads of models [6] spawned, most notably the Active Appearance Models [9, 2, 3].

Dryden & Mardia [7] discriminates landmarks into three subgroups:

- **Anatomical landmarks** Points assigned by an expert that corresponds between organisms in some biologically meaningful way.

- **Mathematical landmarks** Points located on an object according to some mathematical or geometrical property, i.e. high curvature or an extremum point.

- **Pseudo-landmarks** Constructed points on an object either on the outline or between landmarks.

Synonyms for landmarks include homologous points, nodes, vertices, anchor points, fiducial markers, model points, markers, key points et cetera.

A mathematical representation of an n-point shape in k dimensions could be to concatenate each dimension into a kn-vector. The vector representation for planar shapes (i.e. $k = 2$) would then be:

$$\mathbf{x} = [x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n]^{\mathrm{T}} \tag{1}$$

3 Shape Alignment

To obtain a true shape representation – according to our definition – location, scale and rotational effects need to be filtered out. This is carried out by establishing a *coordinate reference* – w.r.t. position, scale and rotation, commonly known as *pose* – to which all shapes are aligned. Some literature also operates with the concept of *pre-shape* as introduced by Kendall [7]. Pre-shape is the last step towards true shape – rotational effects still need to be filtered out.

Below an alignment procedure for obtaining such a coordinate reference is described. This is commonly known as *Procrustes analysis*[2] [1, 5, 7, 11]. This brings the shape set into *shape space*. Adapted to our nomenclature from [7] this is defined as:

> **Definition 3**: *The* **Shape Space** *is the set of all possible shapes of the object in question. Formally, the shape space Σ_k^n is the orbit shape of the non-coincident n point set configurations in the $\mathbb{R}^k$ under the action of the Euclidean similarity transformations.*

But what is the dimension spanned by this shape space? If we have random n point vectors in k Euclidean dimensions the dimensionality is kn. But the alignment procedure peels of dimensionality, i.e. the data now spans only a subspace of kn. The translation removes k dimensions, the uniform scaling one dimension and the rotation $\frac{1}{2}k(k-1)$ dimensions. Thus, the shape space dimensionality is:

$$M = kn - k - 1 - \frac{k(k-1)}{2} \tag{2}$$

If a relationship between the distance in shape space and Euclidean distance in the original plane can be established, the set of shapes actually forms a Riemannian manifold containing the object class in question (e.g. hands). This is also denoted as the *Kendall shape space* [1]. This relationship is called a *shape metric*.

Often used shape metrics include the Hausdorff distance [14], the strain energy [18] and the Procrustes distance [8, 7, 1, 5]. Where the two former compare shapes with unequal amount of points, the latter requires corresponding point sets. However, in the following we will use the celebrated *Procrustes distance*.

[2] As a curiosity Procrustes was the nickname of a robber in Greek mythology called Damastes, who lived by the road from Eleusis to Athens. He offered travelers hospitality on a magical bed that would fit any guest. His humor was to stretch the ones who were too short to fit the bed – until they died – or, if they were too tall, to cut off as much of their limbs as would make them short enough. This rather unpleasant practice continued until Damastes was killed by Theseus, son of Æthra and the Athenian king Ægeus. Another nickname for Damastes was *The one who stretches*.
The term *Procrustes Analysis* was coined by Hurley & Cattell in 1962 [7].

3.1 The Procrustes Shape Distance

The Procrustes distance is a least-squares type shape metric that requires two aligned shapes with one-to-one point correspondence. The alignment part involves four steps:

1. Compute the centroid of each shape.
2. Re-scale each shape to have equal size.
3. Align w.r.t. position the two shapes at their centroids.
4. Align w.r.t. orientation by rotation.

Now, the squared Procrustes distance between two shapes, $\mathbf{x}_1$ and $\mathbf{x}_2$, is simply the sum of the squared point distances:

$$P_d^2 = \sum_{j=1}^{n} [(x_{j_1} - x_{j_2})^2 + (y_{j_1} - y_{j_2})^2] \tag{3}$$

Since the *centroid* of a shape is the center of mass of the physical system consisting of unit masses at each landmark, this is easily calculated as:

$$(\overline{x}, \overline{y}) = \left(\frac{1}{n} \sum_{j=1}^{n} x_j, \frac{1}{n} \sum_{j=1}^{n} y_j \right) \tag{4}$$

To perform step two we obviously need to establish a *shape size metric*:

Definition 4: *A* **shape size metric** $\mathrm{S}(\mathbf{x})$ *is any positive real valued function of the shape vector that fulfils the following property (like all other metrics):*

$$\mathrm{S}(a\mathbf{x}) = a\mathrm{S}(\mathbf{x})$$

In the following the *Frobenius norm* (or 2-norm) is used as a shape size metric:

$$\mathrm{S}(\mathbf{x}) = \sqrt{\sum_{j=1}^{n} [(x_j - \overline{x})^2 + (y_j - \overline{y})^2]} \tag{5}$$

Another often used scale metric is the *centroid size*:

$$\mathrm{S}(\mathbf{x}) = \sum_{j=1}^{n} \sqrt{(x_j - \overline{x})^2 + (y_j - \overline{y})^2} \tag{6}$$

This metric also posses the interesting property that $2n\mathrm{S}(\mathbf{x})^2$ equals the sum of the inter-landmark distances [7].

Turning to step three a Singular Value Decomposition (SVD) is applied [1]:

1. Arrange the size and position aligned $\mathbf{x}_1$ and $\mathbf{x}_2$ as $n \times k$ matrices (in the planar case $k = 2$).
2. Calculate the SVD, $\mathbf{UDV}^{\mathrm{T}}$, of $\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2$ in order to maximize the correlation between the two sets of landmarks.
3. The rotation matrix needed to optimally superimpose $\mathbf{x}_1$ upon $\mathbf{x}_2$ is then $\mathbf{VU}^{\mathrm{T}}$:

$$\mathbf{VU}^{\mathrm{T}} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{7}$$

3.2 Generalized Procrustes Analysis

All though an analytic solution exists [12, 7] to the alignment of a set of planar shapes the following iterative approach to generalized Procrustes analysis [1, 5] will suffice.

1. Choose an initial estimate of the mean shape (e.g. the first shape in the set).
2. Align all the remaining shapes to the mean shape.
3. Re-calculate the estimate of the mean from the aligned shapes.
4. If the estimated mean has changed return to step 2.

Convergence is thus declared when the mean shape does not change significantly within an iteration. Bookstein [1] notes that two iterations of the above should be sufficient in most cases.

The remaining question is how to obtain an estimate of the mean shape (or shape prototype). The most frequently used is the *Procrustes mean shape* or just the *Procrustes mean* (also referred to as the Frechét mean). Let N denote the number of shapes, then the Procrustes mean is:

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{8}$$

Further, to avoid any shrinking or drifting of the mean shape – size and orientation should be properly fixed at each iteration by normalization.

Now, lets return to the concept of a *manifold* introduced earlier. Due to the size normalization all shape vectors live on a subpart of an kn-dimensional hyper sphere. This hyper sphere represents *all* shapes of k landmarks in n Euclidean dimensions, but the subpart in question represents only shapes of the object class in question, e.g. hand outlines. This subpart constitutes the so-called *manifold*.

3.3 Tangent Space Projection

It is a fact that linear methods are nice. Well-behaved, have good performance where applicable and very well understood. However, these stand in striking contrast to the curved surface of high dimensional produced by the Procrustes analysis. This is the (partial) aim of a tangent space projection. Modifying the shape vectors to form a hyper plane, instead of a subpart of a hyper sphere. Further, the Euclidean distance in this plane can be employed as shape metric instead of the true geodesic distance, i.e at the hyper sphere surface.

In fig. 2 (left) a shape vector, $\mathbf{x}$, the mean shape, $\overline{\mathbf{x}}$, and the tangent space (at the pole defined by the mean shape) is shown in a planar projection. The remaining two sketches in fig. 2 show two tangent space projections of which we will focus on the former.

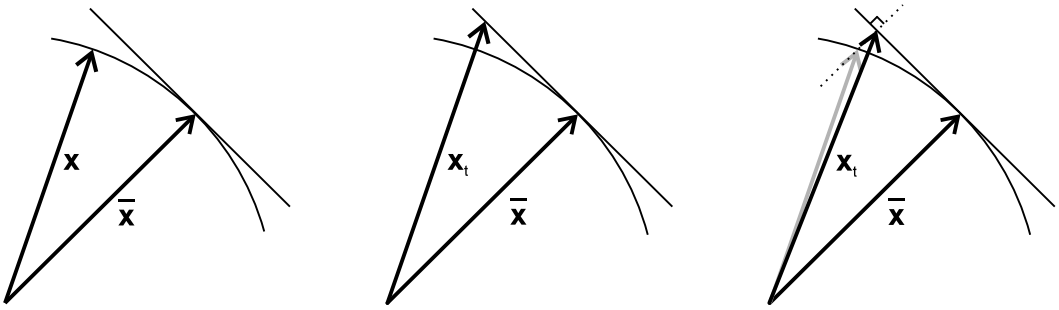


Figure 2: Left: One shape vector drawn from a population, $\mathbf{x}$ and the mean, $\overline{\mathbf{x}}$. Middle: Tangent space projection by scaling. Right: Tangent space projection by scaling and shape modification.

This tangent space projection linearizes by scaling. From fig. 2 (middle) we see that $\mathbf{x}_t$'s projection onto $\overline{\mathbf{x}}$ is $\overline{\mathbf{x}}$ i.e.:

$$\overline{\mathbf{x}} = \frac{\overline{\mathbf{x}} \cdot \mathbf{x}_t}{|\overline{\mathbf{x}}|^2}\overline{\mathbf{x}} = \beta\overline{\mathbf{x}} \tag{9}$$

Substituting $\mathbf{x}_t$ with $\alpha\mathbf{x}$ in the scaling factor, β, we get:

$$\beta = 1 = \frac{\overline{\mathbf{x}} \cdot \mathbf{x}_t}{|\overline{\mathbf{x}}|^2} = \alpha\frac{\overline{\mathbf{x}} \cdot \mathbf{x}}{|\overline{\mathbf{x}}|^2} \tag{10}$$

And finally

$$\mathbf{x}_t = \alpha\mathbf{x} = \frac{|\overline{\mathbf{x}}|^2}{\overline{\mathbf{x}} \cdot \mathbf{x}}\mathbf{x} \tag{11}$$

This type of tangent space projection is carried out on a 2D projection of four and four hundred shape vectors in fig. 3.

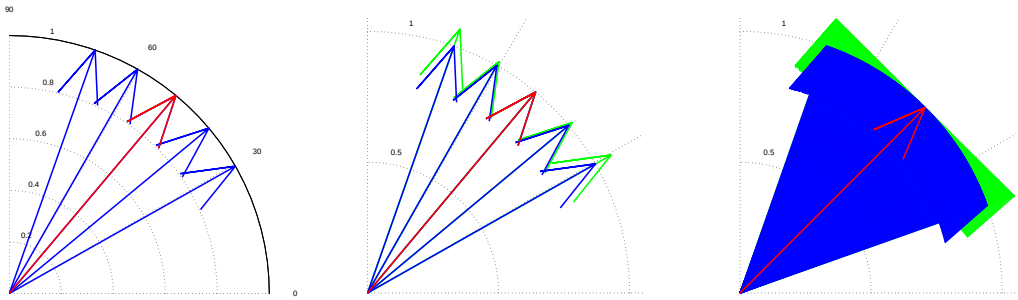


Figure 3: Left: A planar projection of four aligned shapes (mean shape shown in red). Middle: Same as left with tangent space projection (shown in green). Right: Same as middle on four hundred vectors.

To illustrate the impact of this technique consider the toy example containing a set of rectangles in fig. 4 (left). Besides non-shape information (i.e. scale, rotation and translation) these were generated by varying one parameter only, namely the aspect ratio. Due to the tangent space projection all non-linearities induced by the Procrustes analysis in fig. 4 (middle) are removed in fig. 4 (right). As we shall see later this obviously affect the results when trying to model shape variation using linear methods.

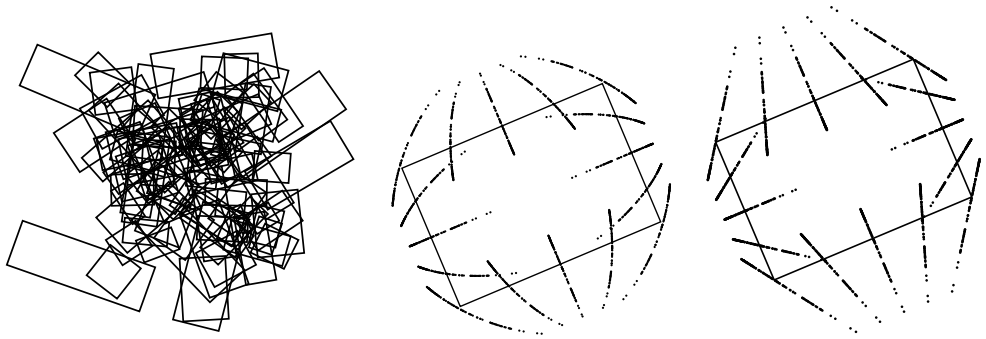


Figure 4: Left: Set of 100 unaligned artificially generated rectangles containing 16 points each. Middle: Mean shape and point cloud from Procrustes aligned rectangles. Right: Mean shape and point cloud from Procrustes aligned rectangles projected into tangent space.

4 Modeling Shape Variation

Whereas the previous section brought the set of shapes into a frame of reference, this section presents a popular method to describe the variation within that frame. This is also known as *shape decomposition*. The application of such are numerous: domain specific interpretations/data exploration, simulation, compression, prediction, outlier detection et cetera.

Though this presentation only deals with one particular decomposition method appealing alternatives exist, e.g. Maximum Autocorrelation Factors (MAF) [16].

4.1 Principal Component Analysis

This section aims at being a self-contained presentation of how the Principal Component Analysis (PCA) can be derived by means of simple linear algebra and used for modeling shape variation. For at more thorough discussion of this general linear data analysis method refer to a textbook in statistics.

Consider the case of having N planar shapes consisting of n points, where each shape is represented as:

$$\mathbf{x} = [x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n]^{\mathrm{T}} \tag{12}$$

Considering a set covering a certain class of shapes, we will always observe some degree of inter-point correlation. If not, the set either i) contain no variation, or ii) the points are purely random, which implies that the points are not landmarks.

This argumentation leads to the suspicion that there could exist a shape representation accounting for correlation between points. If some point movements were to be (numerically) totally correlated, this could be exploited to reduce dimensionality.

In our case we will seek a linear transformation of the data:

$$\mathbf{y} = \mathbf{M}\mathbf{x} \tag{13}$$

First, consider the mean shape $\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$ and the shape covariance matrix:

$$\mathbf{\Sigma}_{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}} \tag{14}$$

The mean of the $\mathbf{y}$-variables can then be expressed as:

$$\overline{\mathbf{y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i = \frac{1}{N}\sum_{i=1}^{N}\mathbf{M}\mathbf{x}_i = \mathbf{M}\overline{\mathbf{x}} \tag{15}$$

And consequently the covariance of the $\mathbf{y}$'s:

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{y}} &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^{\mathrm{T}} \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{M}\mathbf{x}_i - \mathbf{M}\overline{\mathbf{x}})(\mathbf{M}\mathbf{x}_i - \mathbf{M}\overline{\mathbf{x}})^{\mathrm{T}} \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbf{M}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{M}(\mathbf{x}_i - \overline{\mathbf{x}}))^{\mathrm{T}} \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbf{M}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}}\mathbf{M}^{\mathrm{T}} \\
&= \mathbf{M}\left(\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}}\right)\mathbf{M}^{\mathrm{T}} \\
&= \mathbf{M}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{M}^{\mathrm{T}}
\end{aligned} \tag{16}$$

Then, if we limit ourselves to orthogonal transformations (i.e. $\mathbf{M}^{-1} = \mathbf{M}^{\mathrm{T}}$) left-multiplication by $\mathbf{M}^{\mathrm{T}}$ in (16) yields:

$$\mathbf{M}^{\mathrm{T}}\mathbf{\Sigma_y} = \mathbf{\Sigma_x}\mathbf{M}^{\mathrm{T}} \tag{17}$$

Substitution of $\mathbf{M}^{\mathrm{T}}$ by $\mathbf{\Phi}$ yields:

$$\mathbf{\Sigma_x}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Sigma_y} \tag{18}$$

From (18) it is seen that if $\mathbf{\Phi}$ is chosen as the (column) eigenvectors of the symmetric matrix $\mathbf{\Sigma_x}$, then the covariance of the transformed shapes, $\mathbf{\Sigma_y}$, becomes a diagonal matrix of eigenvalues. In the case of correlated points the smallest eigenvalues will be (close to) zero and the corresponding eigenvectors could be omitted from $\mathbf{\Phi}$, thus reducing the length of $\mathbf{y}$.

In conclusion, to establish a linear transform that de-correlate data vectors, the transformation matrix must be the eigenvectors of the covariance matrix of the original data. In order to back transform from the new set of variables, $\mathbf{y}$, we invert (13), remembering that $\mathbf{M}$ is orthogonal:

$$\mathbf{x} = \mathbf{M}^{-1}\mathbf{y} = \mathbf{M}^{\mathrm{T}}\mathbf{y} = \mathbf{\Phi}\mathbf{y} \tag{19}$$

Typically one would apply PCA on variables with zero mean (notice that the $\mathbf{\Phi}$ is unchanged):

$$\mathbf{y} = \mathbf{M}(\mathbf{x} - \overline{\mathbf{x}}) \qquad , \qquad \mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Phi}\mathbf{x} \tag{20}$$

This celebrated method of dealing with redundancy in multivariate data is known as *Principal Component Analysis* (PCA) or the *Karhunen-Loéve Transform* (KLT). It was introduced back in 1933 by Harold Hotelling [13] based on work by Karl Pearson.

It should be mentioned that the PCA also could be derived as a variance maximizing change of basis given an orthonormality constraint using a Lagrange formulation. Not surprisingly the solution of this constrained optimization problem leads to the same eigenproblem as above.

Returning to fig. 4 PCA on the raw aligned rectangles (middle) decomposes the variation into two components accounting for 99.6% and 0.4% of the total variation, respectively. Knowing that only the aspect ratio changes this seems unsatisfactory.[3] However, linearizing by a tangent space projection, fig. 4 (right), the PCA can decompose all variation using one component.

5 Experiments

To demonstrate the techniques described above, this section provides a case study along with an in-depth discussion of the results. All results should be easily reproduced with Matlab, R, S-Plus or a similar using the annotations available from the internet.[4]

5.1 Data

The foundation for all subsequent analysis is 40 images of human hands record using an Olympus C-2040 digital camera at a resolution of 1600×1200 pixels. Four people contributed with ten images each of their left hand. The first six of each ten were constructed as an equally spaced sequence from maximally to minimally spread fingers. The remaining four were chosen arbitrarily by the test persons under the constraints i) that the palm should face the support and ii) that the hand contour should represent a simple curve (i.e. no crossings).

[3] From a theoretically point of view that is. From a numerically point of view one would discard the small component introduced by the non-linearity.

[4] Annotations, images and data reports may be obtained from `http://www.imm.dtu.dk/~mbs/`.

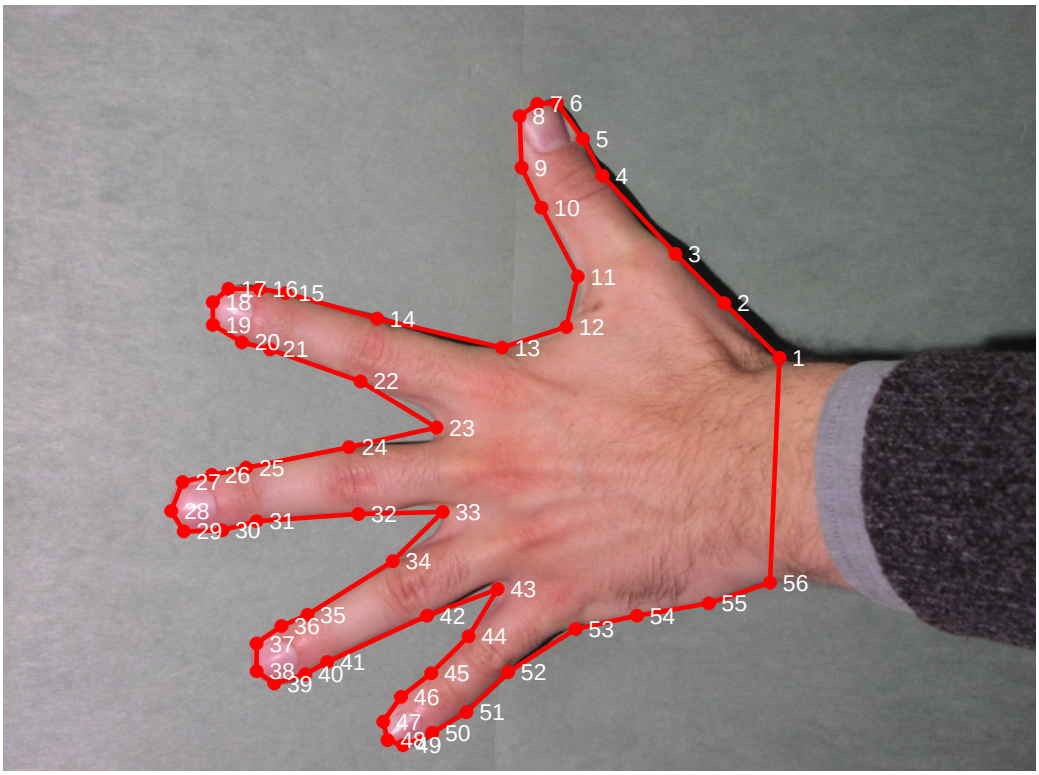


Figure 5: Annotated hand using 56 landmarks.

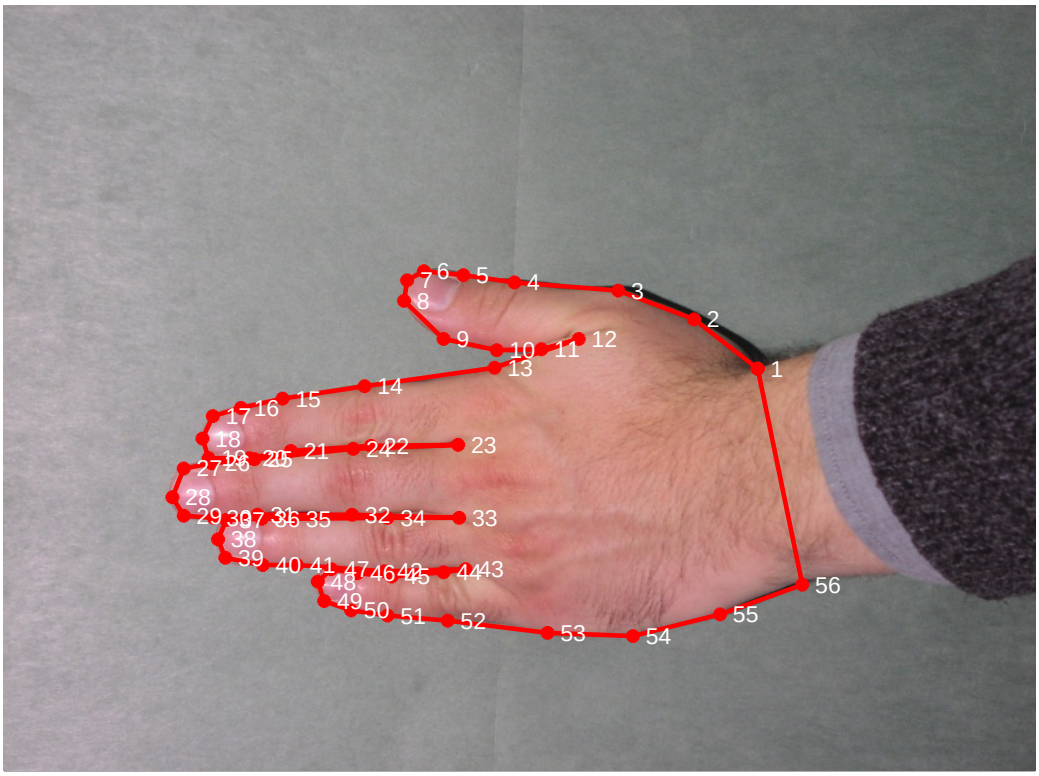


Figure 6: Annotated hand with contracted fingers.

The set of images was brought into correspondence by placing 56 landmarks upon the contour of the hand. An example is given in fig. 5. The distribution of landmarks is as follows:

- 43 type I landmarks (anatomical landmarks).
- 9 type II landmarks (mathematical landmarks): 8, 12, 18, 23, 28, 33, 38, 43, 48.
- 4 type III landmarks (pseudo-landmarks): 2, 11, 54, 55.

The 43 anatomical landmarks were placed i) at the border between forearm and hand, and ii) the knuckles and nails.[5] The four pseudo-landmarks were placed (also by hand) to minimize the distance from the linear spline defined by the 56 landmarks and the actual hand contour.

5.1.1 Remarks regarding the annotation

Specifically regarding hands, fig. 6 indicates that the landmarks 11, 12 are slightly problematic and must be placed with caution. There is further a great variability in the visibility of the knuckles, which leads to a higher uncertainty on these landmarks.

In general, one should bear in mind that placing landmarks is per se a subjective task. For optimal accuracy, intra- and inter-grader/annotator variability studies should be carried out (also called a repeatability and reproducibility study). This is accomplished by letting a set of operators annotate the data set several times each. At each annotation the image order should be randomized to remove ordering-bias. From these sets, point error variances between and within operators are estimated and a maximum likelihood set are estimated.

5.2 Alignment

To recover the actual shape variation a Procrustes analysis and a tangent space projection have been carried out on the scatter of the 40 shapes given in fig. 7 (left). The result of this is shown with the Procrustes mean shape superimposed in fig. 7 (right).

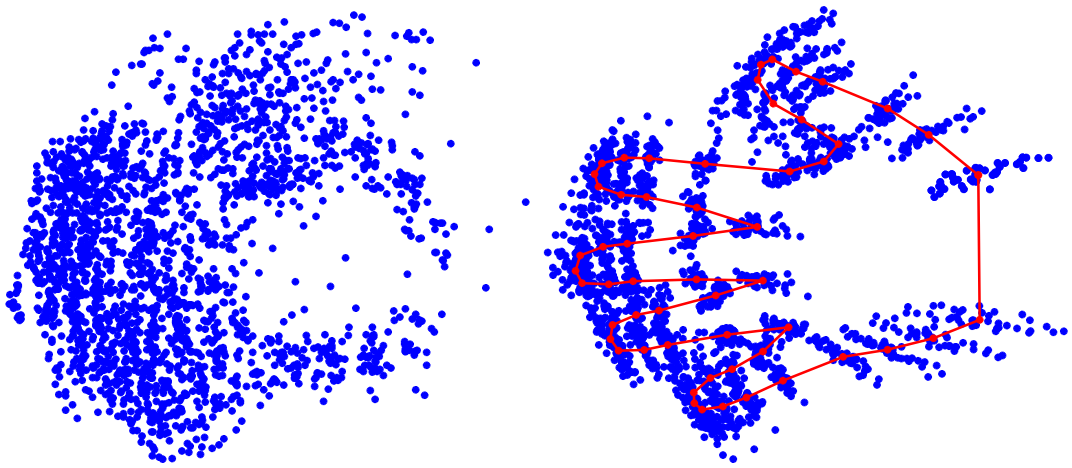


Figure 7: Left: 40 unaligned annotations. Right: 40 aligned annotations with mean shape in red.

At first sight, the point clouds to the right (of the hand) seem fair to model by a multivariate Gaussian. And they are. But true expressions of shape variation they are not. Why? The centroid of the mean shape is roughly between the first and the second knuckle on the middle finger. Now imagine regression lines being calculated for the point clouds 1, 2, 3, 54, 55 and 56. All of these lines would approximately intersect at the centroid. It turns out that this is introduced by the scaling during alignment. Uniform scaling is just moving each shape point along the line spanned

[5]The definition used here is: *Let the line connecting two knuckle landmarks (e.g. 4 and 10) intersect the 'middle' of the knuckle.* This also holds for the landmarks at the start of a nail (e.g. 5 and 9).

by the centroid and the point itself. The reason this occurs is that the contour is denser at the fingertips, which makes the Procrustes analysis emphasize the alignment of this part since the cost is measured 'per-point'. The lesson learned here is that one should be cautious when interpreting these scatter plots.

Moving to the point clouds at the fingertips in fig. 7, interpretation is rendered hopeless due to the overlap. To give a more clear impression Gaussians have been fitted independently to each point cloud in fig. 8 (left). Major and minor axes are three standard deviations. This reveals – not surprisingly for articulated subparts – that the point variation is largest at the fingertips, particularly for the thumb.

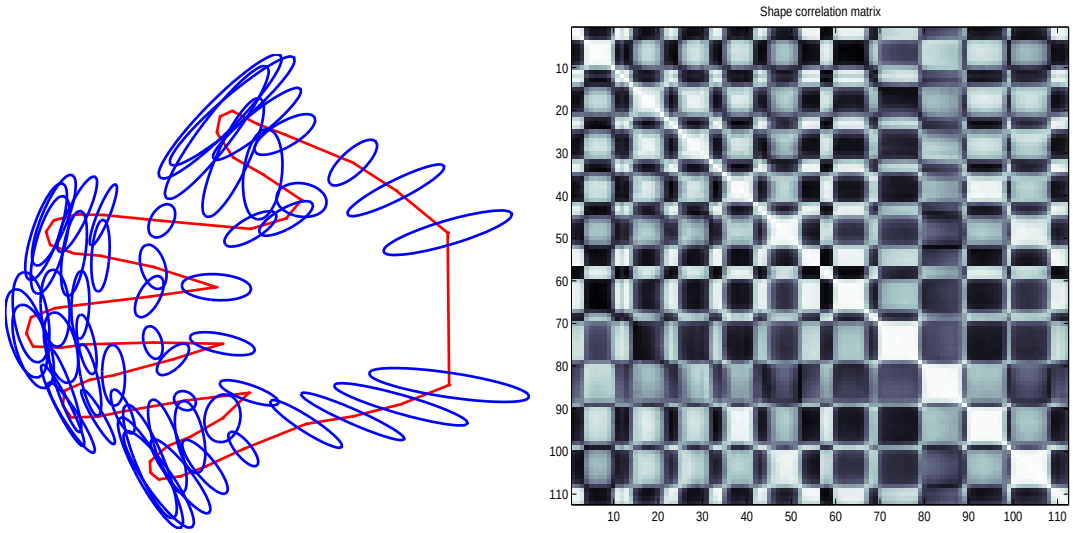


Figure 8: Left: Independent principal component analysis of each model point. Right: Correlation matrix of the annotations (white/grey/black maps to positive/none/negative correlation).

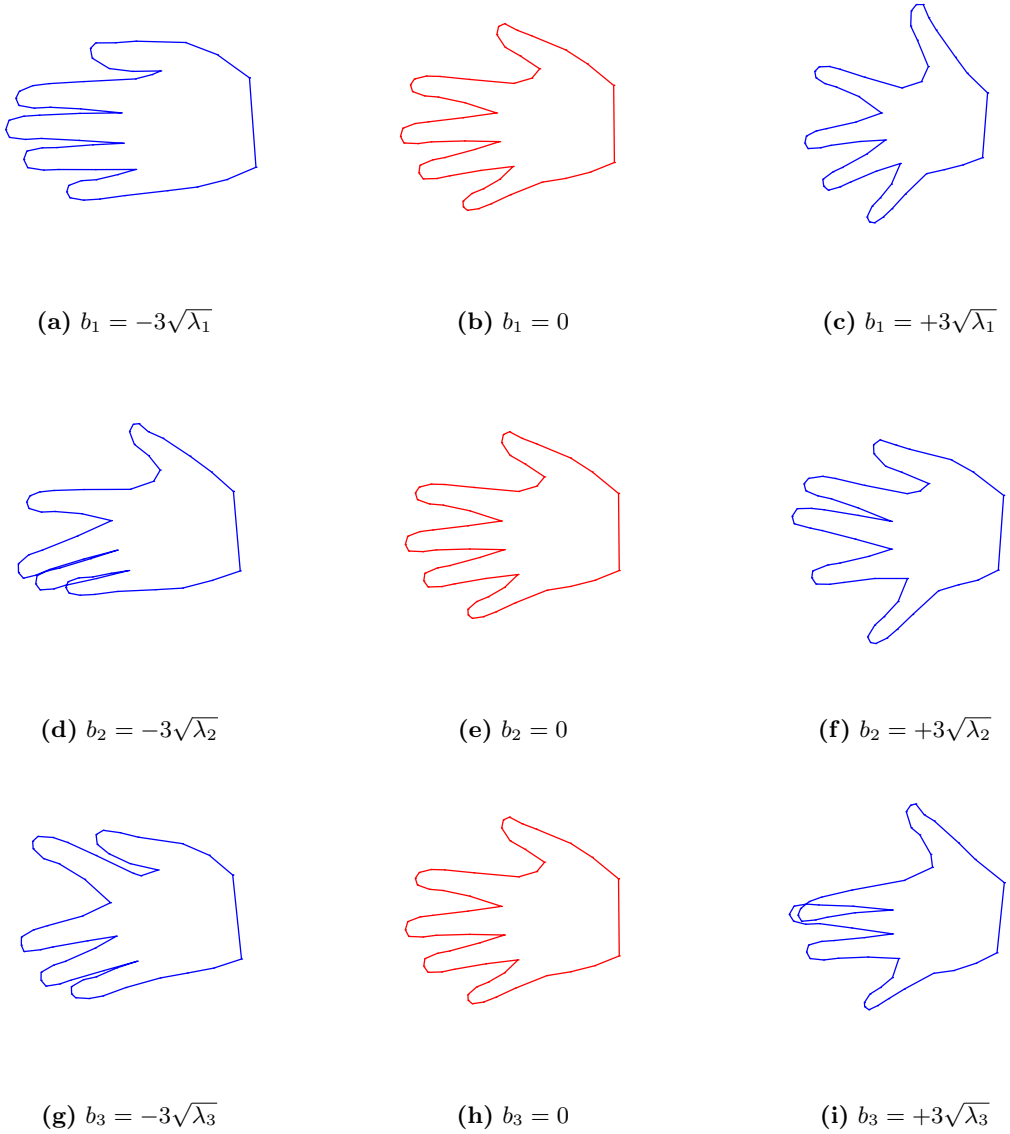
5.3 Statistical Analysis

Hitherto, we have only dealt with isolated point variation. But looking at the correlation matrix in fig. 8 (right) it shows a complex pattern of highly correlated point movements. Starting with the five largest squares at the lower half of the diagonal we have each finger. The index finger e.g. is the square from 70 to 78. Remembering that the 56-point shapes are stored `xxxyyy` we have that this corresponds to the high correlation of the y-positions of points 14–22. This is verified by fig. 5. Another example is the black square 90–98 / 70–78. This verifies that the movement of the index and ring finger is negative correlated in the y-direction. Further, observe that correlations in the x-direction are much less conspicuous, which is no surprise considering the orientation of the mean shape. This emphasizes that the covariance matrix is not invariant to rotations given our shape representation.

A better way to investigate the correlations would be to decompose the set of shapes using the principal component analysis. As mentioned earlier this would result in a an ordered basis where each component is ranked after variance. Modifying one component at a time gives the principal *modes*. For the hand set we have shown mode 1–3 in fig. 9. First mode explains 66% of the total variation in the set. Second and third mode explains 18% and 8% respectively. In total the PCA shape decomposition is able to represent 92% of the variation using only three parameters. For a complete picture refer to fig. 10 (right).

From fig. 9 we can interpret the first mode as how much the fingers are equally spread. Mode two apparently picks up uneven spread fingers, while mode three captures a correlation between i) increasing distance between the middle and index finger, and ii) decreasing distance between the ring and little finger (and vice versa). But also these plots leave us with a warning that one

should be cautious when interpreting. Looking at the top-right plot the hand is smaller than the others. Once again this is simply due to the fact that hands with fingers spread have a larger distance to their centroid and thus are scaled further down in the alignment.



(a) $b_1 = -3\sqrt{\lambda_1}$ **(b)** $b_1 = 0$ **(c)** $b_1 = +3\sqrt{\lambda_1}$

(d) $b_2 = -3\sqrt{\lambda_2}$ **(e)** $b_2 = 0$ **(f)** $b_2 = +3\sqrt{\lambda_2}$

(g) $b_3 = -3\sqrt{\lambda_3}$ **(h)** $b_3 = 0$ **(i)** $b_3 = +3\sqrt{\lambda_3}$

Figure 9: Mean shape deformation using 1st, 2nd and 3rd principal mode, $b_i = -3\sqrt{\lambda_i}$, $b_i = 0$, $b_i = 3\sqrt{\lambda_i}$.

Now, lets tie these principal components back to the data set. Due to the simple linear nature of the decomposition each shape is easily transformed into its principal component equivalent. In fig. 10, a projection of the two first components is given. We see that a Gaussian approximates this distribution to an acceptable degree except for the outliers 38, 40. Looking at the mode plots in fig. 9 we can tell that number 38 has i) a small distance between the index and middle finger, and ii) a large distance between the ring and little finger as opposite for number 40. This is verified in fig. 11. Omitting these two outliers results in capturing 72% variation in the first principal component. This substantial gain has unfortunately been drawn directly from the second and third component as all three now captures 91%.

Finally, we inspect the distribution of the first five principal components – which account for 96% variation – in a scatter plot matrix in fig. 12. Given the relatively small number of observations it should be hard to argue against modeling this set of hands using a multivariate Gaussian (despite the inherent non-linearities).

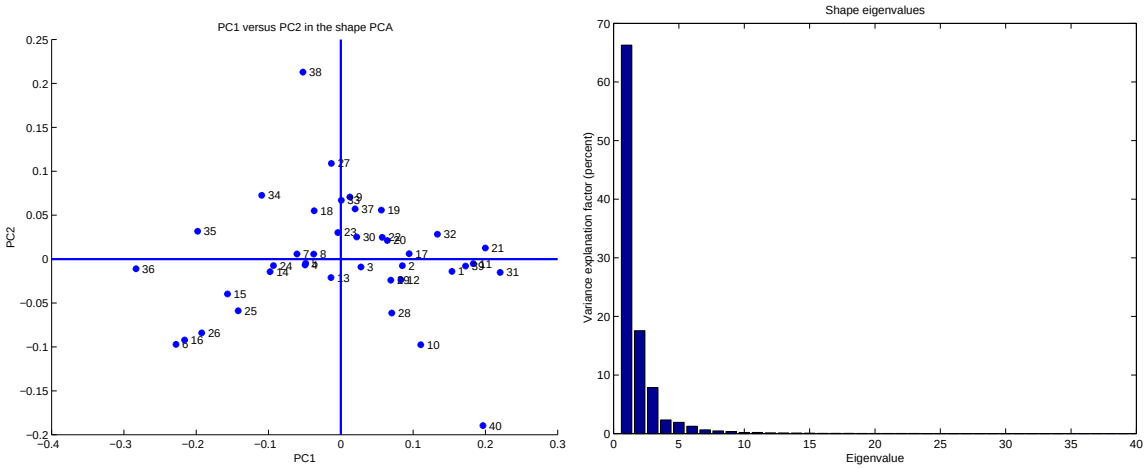


Figure 10: Left: PC1 ($b_{s,1}$) vs. PC2 ($b_{s,2}$) in the shape PCA. Right: Shape eigenvalues in descending order.

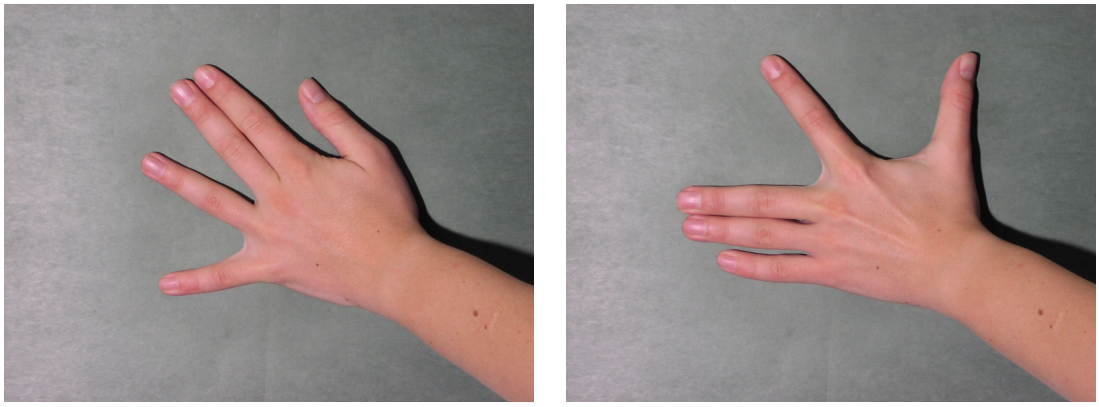


Figure 11: Left: Shape outlier – number 38. Right: Shape outlier – number 40.

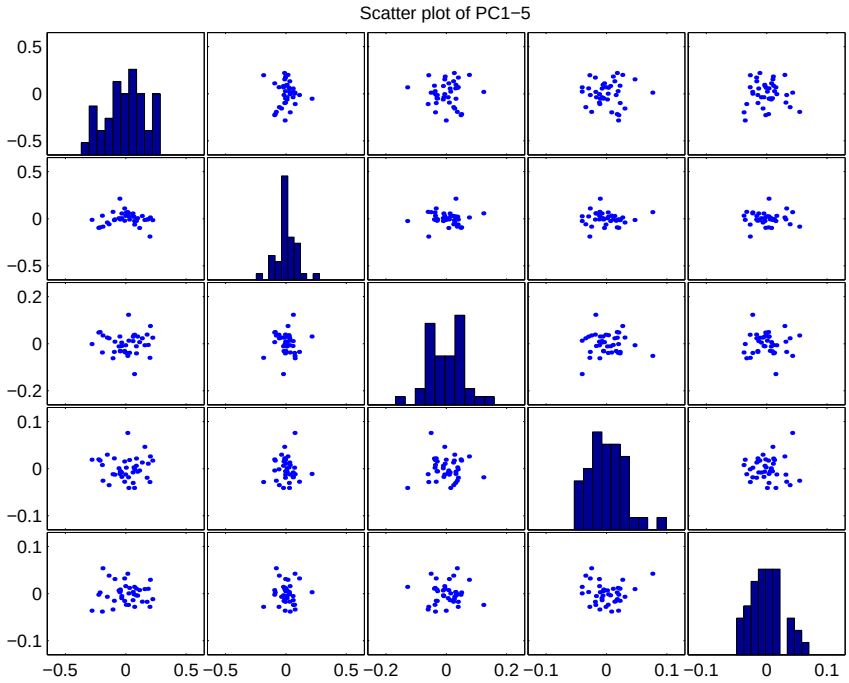


Figure 12: Scatter plot of PC1-5 (PC1 is top-left).

5.4 On the Practical Impact of Tangent Space Projection

Though the results in fig. 4 show promise, the effect of a tangent space projection is rarely that clear. Omitting the tangent space projection in this study leaves imperceptible changes to both mode plots and the structure of the PCA plots. Though very small there is a change in the right direction regarding the compactness of the shape PCA model. PC1 drops from 66.28% to 65.79%, PC2 from 17.56% to 17.51% and using five parameters the accumulated variation are 95.95% and 95.56% with and without tangent space projection respectively. This scenario is typical to most shape models built by the authors. In summary the projection is simple to apply and pleasing from a mathematical point of view the net effect however, is minimal from a numerical point of view.[6]

6 Concluding Remark

We have aimed at serving a brief introduction to the large and intriguing field of statistical shape analysis in an informal and not particular mathematical manner (given the topic). The motivation for doing so is the apparent lack of such an introduction. Techniques have been presented and demonstrated on a data set with well known characteristics, namely hands, and by preparing the data set and three thorough data reports along with this introduction, it is our hope that this will inspire the reader to validate the results and continue the exploration.

As the treatment of the topic is by no mean exhaustive we point to the references for all the exciting details and comments that this text left out.

7 Acknowledgements

The left hands of Panagiotis Karras, Lars Pedersen and the authors are gratefully acknowledged for their participation in the case study.

[6] We stress this is purely our opinion regarding shape modeling settings similar to this.

References

[1] F. L. Bookstein. Landmark methods for forms without landmarks: localizing group differences in outline shape. *Medical Image Analysis*, 1(3):225–244, 1997.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. European Conf. on Computer Vision*, volume 2, pages 484–498. Springer, 1998.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 23(6):681–685, 2001.

[4] T. F. Cootes and Taylor. Active shape models – 'smart snakes'. In *Proc. British Machine Vision Conf., BMVC92*, pages 266–275, 1992.

[5] T. F. Cootes and C. J Taylor. *Statistical Models of Appearance for Computer Vision*. Tech. Report. Oct 2001, University of Manchester, http://www.isbe.man.ac.uk/~bim/, oct 2001.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[7] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998.

[8] N. Duta, A. K. Jain, and M.-P. Dubuisson-Jolly. Learning 2D shape models. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 8–14, 1999.

[9] G.J. Edwards, C. J. Taylor, and T. F. Cootes. Interpreting face images using active appearance models. In *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 300–5. IEEE Comput. Soc, 1998.

[10] Hrafkell Eiriksson. Shape represenation, alignment and decomposition. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, 2001.

[11] C. Goodall. Procrustes methods in the statistical analysis of shape. *Jour. Royal Statistical Society, Series B*, 53:285–339, 1991.

[12] B.K.P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A (Optics and Image Science)*, 4(4):629–42, 1987.

[13] Harold Hotelling. Analysis of complex statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[14] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[15] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. Jour. of Computer Vision*, 8(2):321–331, 1988.

[16] Rasmus Larsen, Hrafnkell Eiriksson, and Mikkel B. Stegmann. Q-MAF shape decomposition. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001, 4th International Conference, Utrecht, The Netherlands*, volume 2208 of *Lecture Notes in Computer Science*, pages 837–844. Springer, 2001.

[17] A. Neumann and C. Lorenz. Statistical shape model based segmentation of medical images. *Computerized Medical Imaging and Graphics*, 22(2):133–143, 1998.

[18] S. Sclaroff and A. P. Pentland. Modal matching for correspondence and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):545–61, 1995.

[19] D'Arcy W. Thompson. *On Growth and Form*. Cambridge University Press, 1917.