

Q-MAF Shape Decomposition

Rasmus Larsen, Hrafnkell Eiriksson, and Mikkel B. Stegmann

Informatics and Mathematical Modelling, Technical University of Denmark
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark
{rl, he, mbs}@imm.dtu.dk, <http://www.imm.dtu.dk/>

Abstract. This paper address the problems of generating a low dimensional representation of the shape variation present in a set of shapes represented by a number of landmark points. First, we will present alternatives to the featured Least-Squares Procrustes alignment based on the L_∞ -norm and the L_1 -norm. Second, we will define a new shape decomposition based on the Maximum Autocorrelation Factor (MAF) analysis, and investigate and compare its properties to the Principal Components Analysis (PCA). It is shown that Molgedey-Schuster algorithm for Independent Component Analysis (ICA) is equivalent to the MAF analysis. The shape MAF analysis utilises the natural order of landmark points along shape contours.

1 Introduction

The Point Distribution Model (PDM) based on PCA of the Active Shape Model (ASM) [1,2] has been succesfully applied to the modelling of the shape of biological objects based on training sets represented by corresponding points. For use for simulation, prediction, or segmentation [3,4,1] a fair number of landmarks are necessary in order to achieve sufficiently good or realistic models. Landmarks are often distributed along outlines or on surfaces equi-distantly or using some other scheme.

Given their representation the objects are aligned wrt. translation, rotation, and scale (e.g. bmo. a Procrustes analysis [5]). Finally, the residual variation is decomposed into latent variables and a low dimensional representation is obtained by retaining only the most important of these. The decomposition of the variation has been based on a number of transformations, most importantly PCA [1,4]. The use of Fourier modes and wavelets are also reported [6,7].

We propose that the choice of landmarks carries important implications for the alignment when using the L_2 -norm based Procrustes analysis and thereby for the resulting shape models. In order to lessen this effect we investigate the use of the L_∞ -norm and the L_1 -norm for aligning the shapes.

Furthermore, we propose an extension to the PCA PDM using the MAF analysis. The MAF by Paul Switzer [8] analysis was originally proposed as an alternative transformation of multivariate spatial imagery to the celebrated PCA transform. In the MAF analysis we seek a transformation that maximises the

autocorrelation between neighbouring observations (e.g. pixels). The basic assumption of the MAF analysis is that the interesting signal exhibits high autocorrelation, whereas the noise exhibits low autocorrelation. By building the additional information of the structure of the observations into the model, application examples (cf. [9,10]) result in a better separation between signal components in one end of the eigenvalue spectrum and noise components in the other end. This is particularly the case when some noise components have higher variance than some signal components. Because the PCA PDM is based on the first say t modes of variation MAF will in this case result in better models than PCA.

The MAF analysis requires estimation of the covariance matrix of the data as well as the covariance matrix of the difference between the original data and a spatially shifted version of the data. Preliminary results are presented in [11]. In an Appendix A we will show that the Molgedey-Schuster algorithm for ICA is equivalent to MAF analysis.

2 Metacarpal Data Set

The proposed methods are illustrated on annotated outlines of 24 metacarpals. The annotations are based on 2-D wrist radiographs, an example is shown in the background of Fig. 3(c). The annotations are prone to errors in the distal and especially in the proximal ends due to the bones being overlaid in the 2-D projection of the radiograph and thus difficult to discern. The annotation variation is therefore point dependent. We believe this to be a common problem to manual annotation. In order to be able to quantify this in the analyses we have simulated two metacarpal datasets. These datasets are generated by first fitting a B-spline to each of the metacarpal outlines after alignment, and second by adding independent Gaussian noise. The noise variance varies smoothly along the contours. It is small along the sides of the metacarpals and high at the ends.

3 Shape Alignment and Choice of Landmarks

Let there be given p training examples for a given shape class, and let each example be represented by a set of n landmark points (x_{ij}, y_{ij}) , $i = 1, \dots, p$ and $j = 1, \dots, n$. The alignment problem in 2D consists of estimating an average shape, $\boldsymbol{\mu}$, and pose parameters for each shape. Let the pose parameters be scale: $\beta_i \in \mathbb{R}_+$, rotation: $\psi \in [0; 2\pi[$, and translation: $\boldsymbol{\gamma}_i \in \mathbb{R}^2$. Then using a multiple linear regression formulation as described in [5] the alignment problem consists of a minimisation of a vector function

$$\mathbf{F} = \begin{bmatrix} \boldsymbol{\mu} - \mathbf{Z}_1 \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\mu} - \mathbf{Z}_{p-1} \boldsymbol{\theta}_{p-1} \\ \boldsymbol{\mu} - \mathbf{Z}_p [1 \ 0 \ 0 \ 0]^T \end{bmatrix}, \text{ where } \mathbf{Z}_i = \begin{bmatrix} x_{i1} & -y_{i1} & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x_{in} & -y_{in} & 1 & 0 \\ y_{i1} & x_{i1} & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ y_{in} & x_{in} & 0 & 1 \end{bmatrix},$$

wrt. $\theta_i = [\beta_i \cos \psi_i, \beta_i \sin \psi_i, \gamma_i^T]^T$. Note, that the average shape is constrained wrt. to size, rotation, and translation by alignment with the last shape. The i th aligned shape is given by $\mathbf{Z}_i \theta_i$. Generalised Procrustes analysis is obtained by minimising the L_2 -norm of this vector function. We will investigate the use of the L_1 -norm and the L_∞ -norm.

Different choices of landmarks for a given class of objects will result in differences in alignment. Consider a set of 10 triangles generated by adding i.i.d. Gaussian noise to an equi-lateral prototype. We will represent these triangles by (1) the coordinates of the corner points, and (2) the coordinates of the corner points and 19 points placed equi-distantly on the lower side. The triangles are shown in Fig. 1(a). The L_2 Procrustes alignment results in the aligned shapes shown in Figs. 1(b) and 1(e). Here the dense sampling of the lower side results in more emphasis being put on the alignment of this side. In the aligned dataset the variation of the points on the densely sampled side exhibit smaller variation, whereas the third corner point exhibit larger variation than we observed in the 3 point representation. By using the L_∞ -norm as shown in Figs. 1(c) and 1(f) we achieve independence of the representation. The L_1 -norm (Figs. 1(d) and 1(d)) increases the effect of putting emphasis on densely sampled shape segments. The L_1 -norm on the other hand aligns the lower side perfectly, and places all variation on the top point. In effect the top point is regarded (wrongly) as an outlier.

4 Maximum Autocorrelation Factor Analysis

Let the spatial (or temporal) covariance function of a multivariate stochastic variable, \mathbf{Y}_k , where k denotes spatial position and Δ a spatial shift, be $\Pi(\Delta) = \text{Cov}\{\mathbf{Y}_k, \mathbf{Y}_{k+\Delta}\}$. Then by letting the covariance matrix of \mathbf{Y}_k be Σ and defining the covariance matrix of the difference process $\Sigma_\Delta = \text{Cov}\{\mathbf{Y}_k - \mathbf{Y}_{k+\Delta}\}$, we find

$$\Sigma_\Delta = 2\Sigma - \Pi(\Delta) - \Pi(-\Delta) \quad (1)$$

Then the autocorrelation in shift Δ of a linear combination of \mathbf{Y}_k is

$$\text{Corr}\{\mathbf{w}_i^T \mathbf{Y}_k, \mathbf{w}_i^T \mathbf{Y}_{k+\Delta}\} = 1 - \frac{1}{2} \frac{\mathbf{w}_i^T \Sigma_\Delta \mathbf{w}_i}{\mathbf{w}_i^T \Sigma \mathbf{w}_i}. \quad (2)$$

This quantity is maximised by minimisation of the Rayleigh coefficient in the second term. Therefore, the MAF transform is given by the set of conjugate eigenvectors of Σ_Δ wrt. Σ , $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_m]$, corresponding to the eigenvalues $\kappa_1 \leq \dots \leq \kappa_m$ [8]. The resulting new variables are ordered so that the first MAF is the linear combination that exhibits maximum autocorrelation. The i th MAF is the linear combination that exhibits the highest autocorrelation subject to it being uncorrelated to the previous MAFs. The autocorrelation of the i th component is $1 - \frac{1}{2}\kappa_i$.

Let the tangent space coordinates of the aligned shapes with the origin placed at the mean shape be the rows of a data matrix, \mathbf{X} . The PCA decomposition

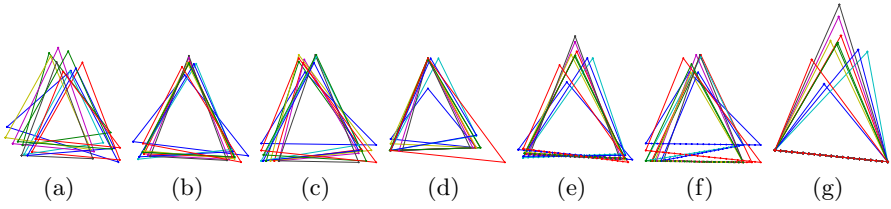


Fig. 1. (a) 10 unaligned triangles, (b-d) alignment based on the L_2 , L_∞ , and the L_1 norms, respectively. (e-g) aligned as (b-d) but with an additional 19 landmarks included distributed equidistantly on the lower side.

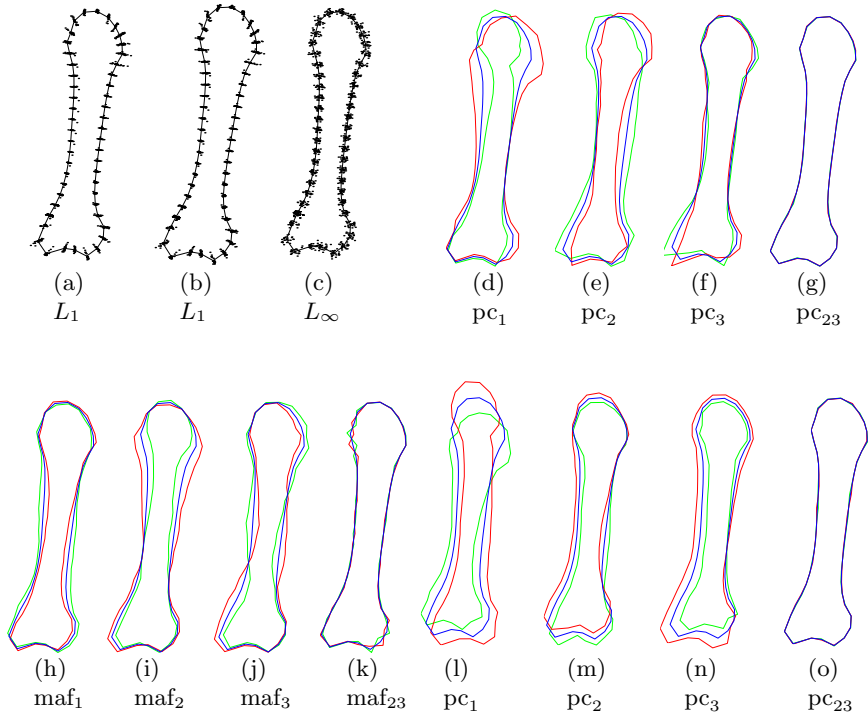


Fig. 2. (a)-(c) The metacarpal dataset aligned and projected into tangent space using the L_1 -norm, L_2 -norm, and L_∞ -norm, respectively. (d)-(g) Principal components of the tangent space coordinates aligned using the L_2 -norm. (h)-(k) Maximum autocorrelation factors of the tangent space coordinates aligned using the L_2 -norm. (l)-(o) Principal components of the tangent space coordinates aligned using the L_∞ -norm. The blue curve is the mean shape the red and green curves correspond to ± 5 standard deviations. The distal end is up and the proximal end is down.

of the dataset can be implemented by extraction of the right singular vectors of a singular value decomposition of \mathbf{X} (cf. [12]). These eigen vectors can also

be extracted by an eigenvalue decomposition of the symmetric matrix $\mathbf{X}^T \mathbf{X}$. In statistical terms this is called an R-mode analysis.

Now, let \mathbf{X}_Δ be a matrix where the x and the y coordinates of each row of \mathbf{X} have been cyclicly rotated Δ shifts to the right. Consider a stochastic variable consisting of the coordinates of a point across the shape training examples, observations of which are given by the columns of \mathbf{X} . Then estimators for the covariance matrices in the MAF eigen problem in Eq. (2) are given by¹

$$\hat{\Sigma} = \frac{1}{2n} \mathbf{X} \mathbf{X}^T \quad \hat{\Sigma}_\Delta = \frac{1}{2n} (\mathbf{X} - \mathbf{X}_\Delta)(\mathbf{X} - \mathbf{X}_\Delta)^T \quad (3)$$

It can be noted that we solve the MAF problem in Q-mode. Therefore we will denote the resulting transform Shape Q-MAF. The modes of variation from the average shape are given by $\mathbf{X}^T \mathbf{W}$. The low number of MAF components will exhibit variations where neighbouring landmark points deform similarly.

5 Analysis of the Metacarpal Data Set

The alignment of the dataset consisting of 24 annotated metacarpals using three different norms is shown in Figs. 2(a)–2(c). The alignments are computed using an optimisation software described in [13].

In Figs. 2(d)–2(g) and 2(h)–2(k), the variations of four (3 first and the last) principal components and Q-MAFs are shown. It turns out that the Q-MAF modes constitute a decomposition of (localised) spatial frequency along the contour with frequency increasing with mode number. Furthermore, the first two modes are easily interpreted as thickness of the cortical bone/aspect ratio, mode three as bending. In the high order number modes variations composed of neighbouring points deforming in opposite directions are concentrated.

The PCA eigen modes are less easily interpreted and it seems that many low number modes are devoted to descriptions of variations of the proximal end. These are variations that may partially stem from annotation arbitrariness. We would normally expect the most important variation in a dataset like this to be variations in aspect ratio. However, because the L_2 -norm alignment has resulted in all shapes having the same length, this effect is transformed into thickness variation, which may lead to mis-interpretations. In Figs. 2(l)–2(o) 4 principal components calculated from the dataset aligned using the L_∞ -norm are shown. Where the L_2 assigns high weight to alignment of the ends of the bone (i.e. favoring many small deviations on the sides to larger deviations in the ends), the L_∞ norm is insensitive to the sampling density. We therefore in some sense arrive at a more natural alignment. As we would expect, the first mode is aspect ratio varying from long and slender to short and stocky.

The more general variant of ASM, Active Appearance Models (AAM) [14] is based upon a PCA decomposition of a set of L_2 Procrustes aligned shapes. An AAM also includes a pixel-wise texture model. This is built by sampling object pixel intensities in each training example followed by a PCA. Consequently,

¹ The variables are corrected for ensemble means so $\hat{\Sigma}$ is a sums-of-squares matrix.

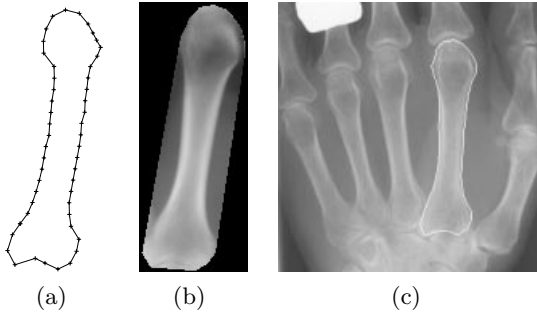


Fig. 3. (a) Metacarpal annotation. (b) Synthesised AAM metacarpal image. (c) AAM segmentation result.

AAMs can synthesise images of the object class in question, see Fig. 3(b). For segmentation purposes the model parameters are adjusted until the synthesised image best matches the unknown image, i.e. under some similarity measure.

Upon the 24 metacarpal shapes and their corresponding radiographs, a leave-one-out performance analysis has been carried out. Texture models consisted of $\sim 10\,000$ pixels. The AAM was initialised using a search-based initialisation method [15], that failed once in each of the three models due to an overlapping metacarpal-1. These were discarded prior to assessment. An example segmentation is shown in Fig. 3(c). The average point to border distance for the L_1 -, L_2 -, and L_∞ -norm AAMs were $1.06 \pm .25$, $1.04 \pm .21$ and $1.09 \pm .25$ pixels, respectively.

In order to evaluate the use of PCA vs. MAF and of the L_2 -norm vs. the L_∞ -norm for building shape models we have tested on the simulated dataset. The models are trained on 24 shapes, and validated on another 24 shapes with independent noise. For all numbers of modes included in the models we have computed the weighted root-mean-square (RMS) error between the fitted model, and the shape without noise. The weights used are the inverse standard deviations of the annotation noise. From plots of the RMS values shown in Fig. 4 we see that the optimal model size is approx. 10, and that the performance is very similar for the tested methods. L_∞ -norm alignment results in RMS errors that are higher than for L_2 -norm alignment.

6 Conclusion

In this paper we have described the use of the infinity norm for shape alignment. This resolves the problem of areas with densely sampled landmarks receiving high weight in the alignment. Furthermore, we have presented the Q-MAF transformation – a novel method for shape decomposition. In applying L_∞ -norm Procrustes alignment as well as the Q-MAF transform to modelling of the outline of metacarpal bones we have achieved better interpretability of the modes of variation at no cost wrt. precision.

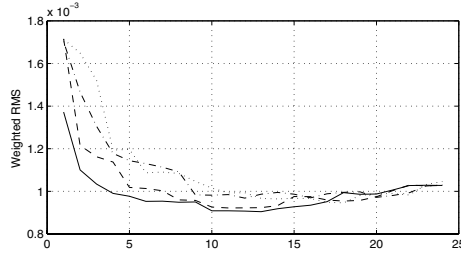


Fig. 4. RMS-error weighed by the annotation noise for an independent validation set. Solid: L_2 , PCA; dashed: L_2 , MAF; dotted: L_∞ , PCA; dash-dotted: L_∞ , MAF.

Acknowledgments

We thank M.D. Lars Hyldstrup, Dept. of Endocrinology, H:S Hvidovre University Hospital, Copenhagen for providing the metacarpal radiographs, and Hans Henrik Thodberg, Ph.D., Pronosco A/S for annotations. Hans Bruun Nielsen, Ph.D., IMM, DTU is thanked for programming B-spline routines.

A Equivalence of ICA and MAF

It turns out that Molgedey-Schusters algorithm for performing ICA [16] is the same as the MAF analysis [8]. Assuming the linear mixing model of independent components analysis $\mathbf{X} = \mathbf{A}\mathbf{S}$, where \mathbf{X} is the $(P \times N)$ data matrix with each row constituting a signal, \mathbf{S} is a matrix of the same form as \mathbf{X} containing independent signals in the rows, and \mathbf{A} is a linear mixing matrix. Furthermore, let \mathbf{X}_Δ and \mathbf{S}_Δ be \mathbf{X} and \mathbf{S} cyclicly shifted Δ steps rowwise. Then the solution is found by forming

$$\mathbf{Q} = \frac{1}{2} [\mathbf{X}_\Delta \mathbf{X}^T + \mathbf{X} \mathbf{X}_\Delta^T] (\mathbf{X} \mathbf{X}^T)^{-1} = \mathbf{A} \left[\frac{1}{2} (\mathbf{S}_\Delta \mathbf{S}^T + \mathbf{S} \mathbf{S}_\Delta^T) (\mathbf{S} \mathbf{S}^T)^{-1} \right] \mathbf{A}^{-1} \quad (4)$$

Due to the independence of the source signals the latter bracketed parenthesis is diagonal. Therefore the mixing matrix can be determined by an eigenvalue decomposition of the matrix \mathbf{Q} , and the source signals up to a scale factor are estimated by $\mathbf{S} = \mathbf{A}^{-1} \mathbf{X}$. Using Eq. (3) we find

$$\mathbf{Q} = \frac{1}{2} [2\mathbf{\Sigma} - \mathbf{\Sigma}_\Delta] \mathbf{\Sigma}^{-1} = \left[\mathbf{I} - \frac{1}{2} \mathbf{\Sigma}_\Delta \mathbf{\Sigma}^{-1} \right]$$

The unity matrix \mathbf{I} has no effect on the eigenvectors, so \mathbf{A} simply consists of the conjugate eigenvectors of $\mathbf{\Sigma}_\Delta$ wrt. $\mathbf{\Sigma}$, i.e. the MAF problem given in Eq. (2).

It is easily shown that the MAF transform is invariant to affine transformations. Therefore we may execute a prewhitening beforehand, thus obtaining $\mathbf{\Sigma} = \mathbf{I}$. Then \mathbf{Q} becomes symmetric yielding $\mathbf{A}^{-1} = \mathbf{A}^T$, and the MAF factors become $\mathbf{A}^T \mathbf{X}_{\text{prewhitened}}$, i.e. the independent components.

References

1. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models – their training and application," *Computer Vision, Graphics and Image Processing*, vol. 61, no. 1, pp. 38–59, Jan. 1995.
2. A. D. Brett and C. J. Taylor, "Construction of 3D shape models of femoral articular cartilage using harmonic maps," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2000*, Scott L. Delp, Anthony M. DiGioia, and Branislav Jaramaz, Eds. 2000, Springer.
3. D. C. Hogg, N. Johnson, R. Morris, D. Buesching, and A. Galata, "Visual modes of interaction," in *2nd International Workshop on Cooperative Distributed Vision, Kyoto, Japan*, 1998.
4. P. R. Andresen, F. L. Bookstein, K. Conradsen, B. K. Ersbøll, J. L. Marsh, and S. Kreiborg, "Surface-bounded growth modeling applied to human mandibles," *IEEE Transactions on Medical Imaging*, vol. 19, no. 11, Nov. 2000, 1053–1063.
5. C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society, Series B*, vol. 53, no. 2, pp. 285–339, 1991.
6. L. H. Staib and J. S. Duncan, "Boundary finding with parametrically deformable models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1061–1075, 1992.
7. A. Neuman and C. Lorenz, "Statistical shape model based segmentation of medical images," *Computerized Medical Imaging and Graphics*, vol. 22, pp. 133–143, 1998.
8. Paul Switzer, "Min/max autocorrelation factors for multivariate spatial imagery," in *Computer Science and Statistics*, L. Billard, Ed. 1985, pp. 13–16, Elsevier Science Publishers B.V. (North Holland).
9. A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, Jan. 1988.
10. A. A. Nielsen and R. Larsen, "Restoration of GERIS data using the maximum noise fractions transform," in *Proceedings of the First International Airborne Remote Sensing Conference and Exhibition*, Strasbourg, France, 1994, vol. 2, pp. 557–568.
11. R. Larsen, "Shape modelling using maximum autocorrelation factors," in *Proceedings of the Scandinavian Image Analysis Conference, SCIA'01, Bergen, Norway, 11–14 June 2001*, 2001, 98–103.
12. R. M. Johnson, "On a theorem stated by Eckart and Young," *Psychometrika*, vol. 28, pp. 259–263, 1963.
13. P. C. Hansen and O. Tingleff, "Robust subroutines for non-linear optimization," Tech. Rep. NI-90-06, Institute for Numerical Analysis, Technical University of Denmark, 1990, 54 pp.
14. T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proceedings of the European Conf. On Computer Vision*. 1998, pp. 484–498, Springer.
15. M. B. Stegmann, R. Fisker, and B. K. Ersbøll, "Extending and applying active appearance models for automated, high precision segmentation in different image modalities," in *Proceedings of the Scandinavian Image Analysis Conference, SCIA'01, Bergen, Norway, 11–14 June 2001*, 2001, 90–97.
16. L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters*, vol. 72, no. 23, pp. 3634–3637, 1994.