# Sparse Modeling of Landmark and Texture Variability using the Orthomax Criterion

Mikkel B. Stegmann, Karl Sjöstrand, Rasmus Larsen

Informatics and Mathematical Modelling, Technical University of Denmark,
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark

## ABSTRACT

In the past decade, statistical shape modeling has been widely popularized in the medical image analysis community. Predominantly, principal component analysis (PCA) has been employed to model biological shape variability. Here, a reparameterization with orthogonal basis vectors is obtained such that the variance of the input data is maximized. This property drives models toward *global* shape deformations and has been highly successful in fitting shape models to new images. However, recent literature has indicated that this uncorrelated basis may be suboptimal for exploratory analyses and disease characterization.

This paper explores the orthomax class of statistical methods for transforming variable loadings into a *simple structure* which is more easily interpreted by favoring sparsity. Further, we introduce these transformations into a particular framework traditionally based on PCA; the Active Appearance Models (AAMs). We note that the orthomax transformations are independent of domain dimensionality (2D/3D etc.) and spatial structure. Decompositions of both shape and texture models are carried out. Further, the issue of component ordering is treated by establishing a set of relevant criteria. Experimental results are given on chest radiographs, magnetic resonance images of the brain, and face images. Since pathologies are typically spatially localized, either with respect to shape or texture, we anticipate many medical applications where sparse parameterizations are preferable to the conventional global PCA approach.

## 1. INTRODUCTION

Due to the frequent noisy and highly complex nature of many medical imaging modalities, constrained solutions are often required. One popular class of constrained image analysis is the various forms of shape models. Here, a top-down approach is taken to the localization of a structure in a medical image using an explicit model of the geometrical layout of the structure supplemented by a set of associated variation patterns. Combined, these two entities should optimally be able to represent the given variability of the structure and nothing apart from that. Hence, only valid solutions can be produced, provided that the model can be fitted with a sufficiently high likelihood. Further, in many applications it may be desirable to be able to extend the use of such models from the classic segmentation or registration scenario, to a level where the model parameterization possesses inherent *interpretive powers* where latent variables are expressed directly. An example of such is disease characterization by surrogate markers, see e.g.[1]

Decomposition of shape and texture variability is predominately carried out by principal component analysis (PCA), which produces a reparameterization with orthogonal basis vectors such that the variance of the input data is maximized. Although this basis is in many senses optimal, recent literature indicate that it might not posses a sufficiently expressive basis for some medical interpretation scenarios.[2–5] Since PCA maximizes variance, new variables (i.e. the principal components) will typically affect the shape or texture globally. In turn, this may lead to confounding of effects due to the chance correlation inherent to limited medical data sets. Interestingly, it has been observed that independent component analysis (ICA) of shape produces new variables showing more localized effects, and thus being able to describe specific pathologies.[2–5] We note that *localization* is often a desirable property for a basis aimed at explaining complex latent relations between pathology and geometry/texture. Consequently, it seems natural to promote transformations favoring locality directly, rather than indirectly. In this paper we explore the orthomax criterion for optimizing sparsity corresponding to new variables being associated to localized modes of variation.

Corresponding author is M. B. Stegmann, E-mail: mbs@imm.dtu.dk, Web: http://www.imm.dtu.dk/∼mbs/.

## 2. RELATED WORK

Although orthomax rotations are well-known within the statistical discipline *factor analysis*, this work is arguably among the first within medical image analysis to explore a method that directly optimizes sparsity. This can be seen as a natural continuation of the previously mentioned work on ICA shape modeling by Üzümcü et al.[2,3] and Suinesiaputra et al.[4,5] in addition to the general literature on alternative parameterizations of shape.

Even for complex biological phenomena, principal component analysis typically yields a very good decomposition of shape variability in a cohort. However, significant non-linearities exist in some cases, which render the implicit assumption of a multivariate Gaussian distribution invalid. Thus, PCA models will yield a poor specificity, leading to potential synthesis of implausible shape configurations. Some of these problematic cases are designed synthetically to emphasize the limitations of a PDM, while others are demonstrating actual, real-world examples of shape variability with dominating non-linearities. Attempts to deal with such non-linearity include the polynomial regression PDM, PRPDM, by Sozou et al.[6] Later, Sozou et al.[6] outperformed this using a back propagation neural network employing a multi-layer perceptron, which resulted in another xPDM acronym; the MLPPDM. A different approach is to employ a kernel-based density estimation of the shape distribution. This was proposed by Cootes and Taylor[7,8] along with a computationally more attractive variant using a Gaussian mixture model to approximate the density function. Building on similar ideas, Heap and Hogg[9] proposed a hierarchical PDM, the HPDM, also based on multiple Gaussian models. Non-linear shape models are also treated in depth by Bowden.[10]

Advances within machine learning that allow working implicitly in infinite dimensional spaces have also been utilized in shape modeling using kernel methods. By employing a variant of non-linear PCA called Kernel PCA (KPCA), complex non-linear shape distributions can be modeled. This was demonstrated on shapes from projections of varying-angle faces by Romdhani et al.[11] Further developments of this work were presented by Twining and Taylor[12] on synthetic shapes, and shapes from images of nematode worms.

While PCA decomposes variation by maximization of variance, other measures may also be of interest when a shape basis is to be chosen. For example, Larsen[13,14] and Larsen et al.[15] chose to maximize autocorrelation along 2D shape contours using the maximum autocorrelation factors (MAF) due to Switzer.[16] Hilger and Stegmann[17] later employed MAF as texture basis in Active Appearance Models (AAMs). The MAF approach was further extended to three-dimensional PDMs by Hilger et al.,[18] Larsen and Hilger,[19] and Larsen et al.[20] Interestingly, it turns out that Molgedey-Schusters algorithm for performing ICA[21] is equivalent to MAF analysis, see.[15]

Turning to the specific use of the orthomax criterion, Ramsay and Silverman give an instructive case study on varimax rotation of principal components based on one-dimensional temperature curves.[22] Related to this is also the work by Peterson et al.[23] where two-dimensional contours of the brain structure *corpus callosum* were decomposed using PCA and subsequently rotated using the varimax criterion. While a similar corpus callosum case is presented here, the two papers are contrasted by the depth in which the rotation method is treated, and the depth in which the case study is analyzed, e.g. w.r.t. functional correlates such as IQ, handedness, et cetera.

Chennubhotla and Jepson[24] developed a *sparse PCA* method* bearing resemblance to the original varimax algorithm[25] by employing a sequence of bi-variate rotations. However, rather than optimizing variance, a function composed of the projected data and the basis vectors were investigated. This was carried out with a weight term controlling the transition from a PCA solution to a sparse solution. Examples were given on images, vector fields, and one-dimensional curves.

Regular PCA extracts new variables, the principal components, as linear combinations of the original variables. For interpretation purposes, the problem is that each new variable is a linear combination of *all* original variables. Sparse PCA aims at approximating the properties of regular PCA, while keeping the number of dependent variables, or equivalently, the number of non-zero loadings, small. Recently, Zou and Hastie presented an algorithm for computing sparse loading matrices.[26] It is heavily based on variable selection methods from regression analysis, primarily the elastic net.[27] Similarly to the SCoTLASS[28] method, a constraint is imposed on each loading vector, limiting the sum of absolute loadings. This drives some loadings to exactly zero, producing a sparse loading matrix in the strict sense. Results are given for the classic "pitprops" data set, some simulated data, and the Ramaswamy microarray data set. Results on medical shape data can be found in Ref. 29.

*This method is different from the sparse PCA method by Zou, Hastie and Tibshirani described below.

## 3. METHODS

### 3.1. Principal Component Analysis

Consider a set of $n$ vectors $\{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^p$ having sample dispersion matrix $\mathbf{\Sigma_x}$. These could denote shape given by landmarks, or texture given by image intensities. Principal component analysis (PCA) transforms these vectors into a decorrelated basis $\mathbf{b}$ with dispersion matrix $\mathbf{\Sigma_b} = \text{diag}(\boldsymbol{\lambda})$ by $\mathbf{b} = \mathbf{\Phi}^\mathsf{T}(\mathbf{x} - \overline{\mathbf{x}})$, where $\mathbf{\Phi}$ denotes the eigenvector solution to $\mathbf{\Sigma_x}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Sigma_b}$ and $\overline{\mathbf{x}}$ denotes the sample mean. Each eigenvector holds a variation pattern referred to as a deformation mode, where each of the original $p$ variables is *loaded* by a given amount. Consequently, the terms eigenvectors, deformation modes, and variable loadings will be used interchangeably in the following. Let eigenvalues, $\lambda_i$, and corresponding eigenvectors be ordered so that $\lambda_1 \geq \cdots \geq \lambda_n = 0$ (when $n < p$). The deformation modes given by the higher order part of $\mathbf{b}$ are typically discarded by a variance-based criterion retaining e.g. 95% of trace($\mathbf{\Sigma_b}$) in $k$ modes. A new example in $\mathbb{R}^p$ given by $\mathbf{b}$ can now be synthesized by the projection $\mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Phi}\mathbf{b}$. Examples of using this generative property of PCA for image interpretation include inter-point distance models[30] and the later point distribution models (PDMs).[31]

### 3.2. Sparse Modeling Using the Orthomax Criterion

Orthomax rotations of a principal component basis reintroduce component correlation to obtain a *simple structure* of the final basis. Let $\mathbf{\Phi}$ be a $p \times k$ orthonormal matrix (of column eigenvectors) and $\mathbf{R}$ be an orthonormal rotation matrix in $\mathbb{R}^k$, i.e. $\mathbf{R}^\mathsf{T}\mathbf{R} = \mathbf{I}_k$, where $\mathbf{I}_k$ denotes the $k \times k$ identity matrix. Further, let $\mathbf{R}_{ij}$ denote the scalar element in the $i^{th}$ row and $j^{th}$ column in matrix $\mathbf{R}$. The class of orthomax rotations can now be defined as

$$\mathbf{R}_{orthomax} = \arg\max_{\mathbf{R}} \left( \sum_{j=1}^{k} \sum_{i=1}^{p} (\mathbf{\Phi}\mathbf{R})_{ij}^4 - \frac{\gamma}{p} \sum_{j=1}^{k} \left( \sum_{i=1}^{p} (\mathbf{\Phi}\mathbf{R})_{ij}^2 \right)^2 \right), \tag{1}$$

where $\mathbf{R}_{orthomax}$ denotes the resulting rotation and $\gamma$ denotes the type. This paper investigates $\gamma = 1$ (varimax[25]) and $\gamma = 0$ (quartimax, e.g.[32]). Further rotations include: $\gamma = k/2$ (equamax), and $\gamma = p(k-1)/(p+k-2)$ (parsimax). Orthomax rotations are traditionally computed using a sequence of bi-variate rotations.[25,32] However, since varimax and quartimax are the only cases treated here, this work employ an iterative method based on singular value decomposition (SVD) for solving Equation 1, which is given in Algorithm 1. Notice that this returns the rotated basis, rather than $\mathbf{R}_{orthomax}$. The algorithm is also employed in the statistical language R and the computational system Matlab. It was first described by Horst[33] and independently shortly after in a different – albeit equivalent[34] – formulation by Sherin.[35] The relation between Equation 1 and Algorithm 1 is detailed in Appendix A.

**Algorithm 1** Estimation of Orthomax Rotation for $\gamma \in [0; 1]$

**Require:** $\mathbf{\Phi} \in \mathbb{R}^{p \times k}$, $\gamma$, $q$, $tol$, Diag($\cdot$) (sets off-diagonal elements to zero), $\odot$ Hadamard (element-wise) product

```
R = I_k
d = 0
for i = 1 to q do
    d_old = d
    Λ = ΦR
    [U, S, V] = svd( Φ^T(Λ ⊙ Λ ⊙ Λ − (γ/p) Λ · Diag(Λ^T Λ)) )
    R = UV^T
    d = trace(S)
    if d/d_old < tol then
        break
    end if
end for
Λ = ΦR
return Λ
```

Let us investigate the varimax variation a bit more closely. Let $\mathbf{\Lambda}$ denote the orthomax-rotated basis, $\mathbf{\Phi R}$, and let $\overline{\mathbf{\Lambda}_j^2}$ denote the mean of the $j^{\text{th}}$ column of $\mathbf{\Lambda}$ having its elements squared. From Equation 1 we see that choosing $\gamma = 1$ will yield the maximal variance of the squared rotated variable loadings summed over all modes;

$$p\sum_{j=1}^{k}\left(\frac{1}{p}\sum_{i=1}^{p}(\mathbf{\Lambda}_{ij}^2)^2 - \frac{1}{p^2}\left(\sum_{i=1}^{p}\mathbf{\Lambda}_{ij}^2\right)^2\right) = p\sum_{j=1}^{k}\left(\frac{1}{p}\sum_{i=1}^{p}\left(\mathbf{\Lambda}_{ij}^2 - \overline{\mathbf{\Lambda}_j^2}\right)^2\right). \tag{2}$$

Since $\mathbf{R}$ is an orthonormal matrix, and thus cannot change the squared sum of the new basis vectors in $\mathbf{\Lambda}$, the variance of each column in $\mathbf{\Lambda}$ can only be increased by bringing some variable loadings close to zero, and let others grow large. Hence, a more simple structure of $\mathbf{\Lambda}$ is obtained. This tends to make the components, or the basis vectors, easier to interpret. The cost is that component correlation will be introduced for any rotation of the PCA basis, except for 180 degrees, in which case the variance would remain unchanged. Relaxing $\mathbf{\Phi}$ to be orthogonal, rather than orthonormal, will lead to both non-orthogonal variable loadings (i.e. $\mathbf{\Lambda}^\mathsf{T}\mathbf{\Lambda}$ not diagonal), as well as to correlated variables.[36] It should be added that subgroups of $\mathbf{\Phi}$ can be rotated, while other modes are left unchanged. Thus, dispersions with block diagonals will be obtained. Such subgroups could be determined by identifying clusters in the eigenvalue spectrum of an initial PCA transformation.[37] However, $\mathbf{\Phi}$ will remain orthonormal and all components will be rotated in this paper.

Setting $\gamma = 0$ yields the special case denoted *quartimax*; a method introduced almost simultaneously by several researchers,[32] and which preceded the varimax approach by a few years. In the quartimax case, Equation 1 becomes:

$$\mathbf{R}_{orthomax} = \arg\max_{\mathbf{R}} \sum_{j=1}^{k}\sum_{i=1}^{p}(\mathbf{\Phi R})_{ij}^4. \tag{3}$$

It turns out that this expression minimizes the parsimony criterion put forward by Ferguson (see Ref. 32),

$$\sum_{i=1}^{p}\sum_{j=1}^{k}\sum_{q=1}^{j-1}(\mathbf{\Lambda}_{ij}\mathbf{\Lambda}_{iq})^2, \tag{4}$$

since $\mathbf{R}$ remains orthonormal and therefore does not change the squared sum of loadings. If this sum is squared, then for a single variable, $i$, we have

$$\left(\sum_{j=1}^{k}\mathbf{\Lambda}_{ij}^2\right)^2 = \sum_{j=1}^{k}\mathbf{\Lambda}_{ij}^4 + 2\sum_{j=1}^{k}\sum_{q=1}^{j-1}\mathbf{\Lambda}_{ij}^2\mathbf{\Lambda}_{iq}^2. \tag{5}$$

Consequently, as Equation 5 remains constant when summed over all variables, Equation 4 is minimized when Equation 3 is maximized. In other words, by emphasizing simplicity within rows of $\mathbf{\Lambda}$, quartimax is contrasted to varimax that emphasizes simplicity within columns of $\mathbf{\Lambda}$. Refer to Ref. 32 for further details on the various, but similar, quartimax formulations.

When focusing on shape variability, one important – albeit rare – situation deserves mentioning. Imagine that $k$ is close to $p$. Then $\mathbf{\Lambda}$ will approach the identity matrix, $\mathbf{I}$. This will happen even when the starting point is a very uneven eigenvalue spectrum. Such behavior is of course entirely correct; we should obtain a maximally sparse solution for a set of eigenvectors that span $\mathbb{R}^p$. But the implication for a shape model based on shapes in $\mathbb{R}^d$ ($d = 2$ or $d = 3$ typically) is that the solution depends solely on the original orientation of the $d$-dimensional coordinate system. The solution is in other words not rotation invariant and this fact becomes very apparent when $k$ approaches $p$. In summary, choice of $k$ will greatly influence the level of obtained sparsity, when all modes are rotated. This issue was also commented by Suinesiaputra et al.[4,5]

Obviously, texture models are not affected by the above issue, since $d = 1$. Although the computations in Algorithm 1 becomes substantial when $p$ is very large (say $p = 30000$ for a texture model) the growth is fortunately linear in $p$. Notice that the costly singular value decomposition is carried out on a $k \times k$ matrix, which does not pose a problem, as $k \ll p$ for such models.

Another issue is the ordering of the new variables stemming from an orthomax rotation. To this end, we discuss a set of criteria below that all order components by decreasing value of the criterion.

**Component variance**.
This is the normal ordering of the principal components. Using this criterion, very sparse modes will tend to reside among the last components due to the orthonormality of $\mathbf{\Lambda}$. That is, sparse modes will be scaled more than dense, and consequently lead to smaller component scores.

**Variance of squared loadings**.
As this is the criterion being optimized by the varimax rotation, this ordering may be a natural choice for having sparsity concentrated among the first modes.

**Locality**.
Favorable, if prior knowledge is available regarding interesting subparts of the original $p$-dimensional space. Used by Üzümcü et al.[2,3] and Suinesiaputra et al.[4,5] when ordering sparse, ICA-based shape modes according to their effects along near-circular endo- and epicardial borders in cardiac magnetic resonance images.

**Correlation**.
This ordering is suitable if $k$ has been chosen to produce an appropriate amount of sparsity in the resulting modes and the objective is to find sparse, yet weakly correlated, modes. Those will thus be present in the latter part of the ordered modes.

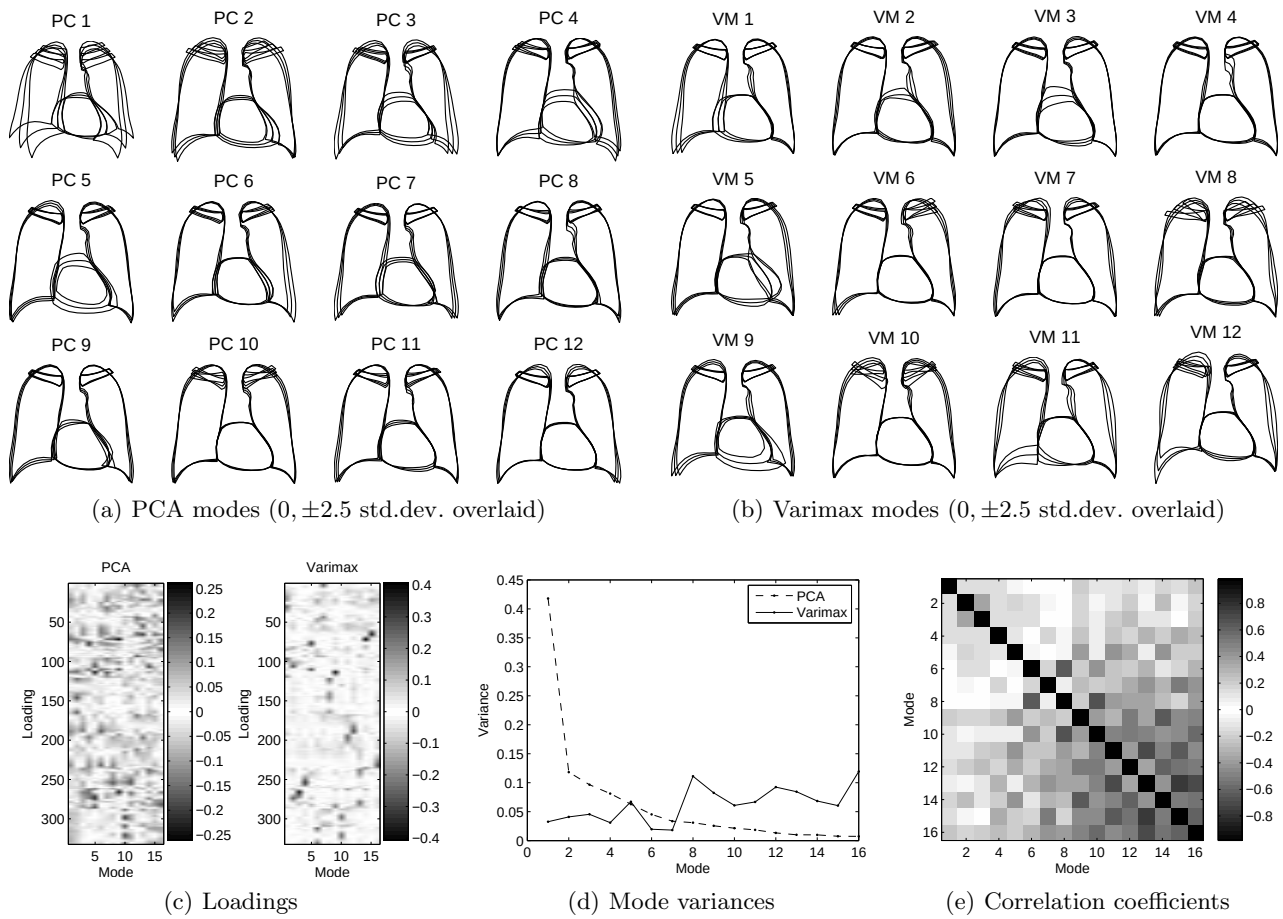
**Autocorrelation**.
Although, sparsity typically is obtained by fairly well-defined coherent parts of the original $p$ dimensional domain, this behavior is not required by design. Ordering by autocorrelation will discriminate between abrupt changes and more smooth coherent modes. However, care has to be taken when estimating the autocorrelation for multi-part or open-contour shapes and for textures.

**Clustering**.
If numerically large variable loadings are localized in several clusters e.g. along a contour in a shape model, then ordering according to the numbers of clusters and cluster size may be interesting.

The following section will demonstrate the use of three of the above criteria.

## 4. EXPERIMENTAL RESULTS

To illustrate the effects of orthomax rotation, three different cases have been selected and decomposed by principal component analysis and subsequently rotated using the varimax criterion. Two shape studies were carried out on contours stemming from chest radiographs and magnetic resonance images (MRI) of the human brain. Normal perspective images of frontal faces formed basis for a case study of sparse texture variability. Varimax-rotated components are compared to principal components in all three cases.




(a) PCA modes $(0, \pm 2.5$ std.dev. overlaid)

(b) Varimax modes $(0, \pm 2.5$ std.dev. overlaid)

(c) Loadings

(d) Mode variances

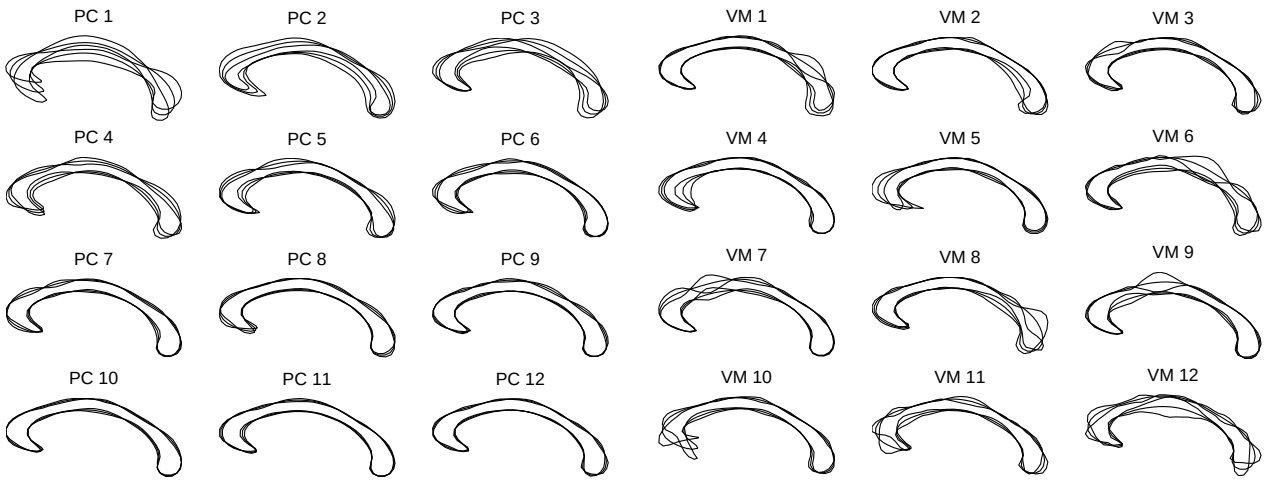
(e) Correlation coefficients

**Figure 1.** Shape modes calculated from 247 chest radiograph annotations of the lungs, heart and clavicles.

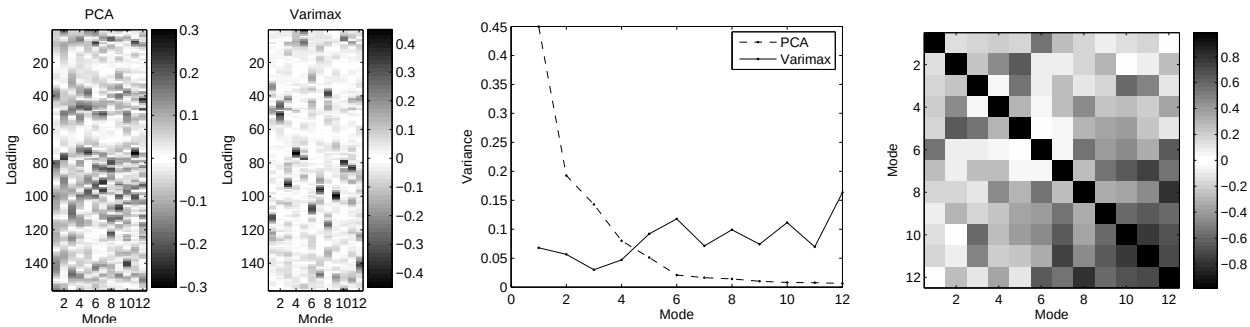
Figure 1 shows a decomposition of 247 chest radiograph annotations of the lungs, heart and clavicles based on 166 landmarks. This data is described in detail in Refs. 38 and 39. In our case, the 16 largest principal components were retained. All varimax modes in Figure 1(b) were sorted by the absolute sum of the correlation coefficients in order to probe for localized yet weakly correlated modes, see Figure 1(e). We see that varimax mode 4 is related to the position of the aortic arch. Modes 3 and 5 relate to heart size, while modes 8 and 10 relate to clavicle orientation. These localized modes are contrasted by the conventional PCA modes shown in Figure 1(a). Another way of visualizing the variable loadings in each case is to relate gray-scales to the magnitudes of the elements in $\mathbf{\Phi}$ and $\mathbf{\Lambda}$. This is carried out in Figure 1(c), which clearly demonstrates the sparsity of the varimax solution. The 'flattening' of the eigenvalue spectrum carried out by the varimax rotation is illustrated in Figure 1(d) where the respective variances are plotted.

Figure 2 shows a decomposition of 62 annotations of the corpus callosum in mid-sagittal brain MRI using 78 landmarks. This data is described in more detail in Refs. 38, 40, and 41. Varimax ordering is similar to the previous case study. In Figure 2(b) we observe that varimax mode 1 relates to the isthmus area, mode 2 to bending of the splenium, mode 3 to the truncus area, while mode 4 is clearly related to the area of the rostrum and genu. In contrast, PCA mode 1 describes a simultaneous bending of the entire corpus callosum with an area change of the rostrum, genu and splenium. Again, the sparsity of the varimax-rotated components can also be appreciated in Figure 2(c).

Figures 3 and 4 show the results of a decomposition of 37 gray-scale face images. Further analyses of this data set can be found in Refs. 42 and 43. Prior to our analyses, all images were compensated for any variation in




(a) PCA modes $(0, \pm 2.5$ std.dev. overlaid) (b) Varimax modes $(0, \pm 2.5$ std.dev. overlaid)



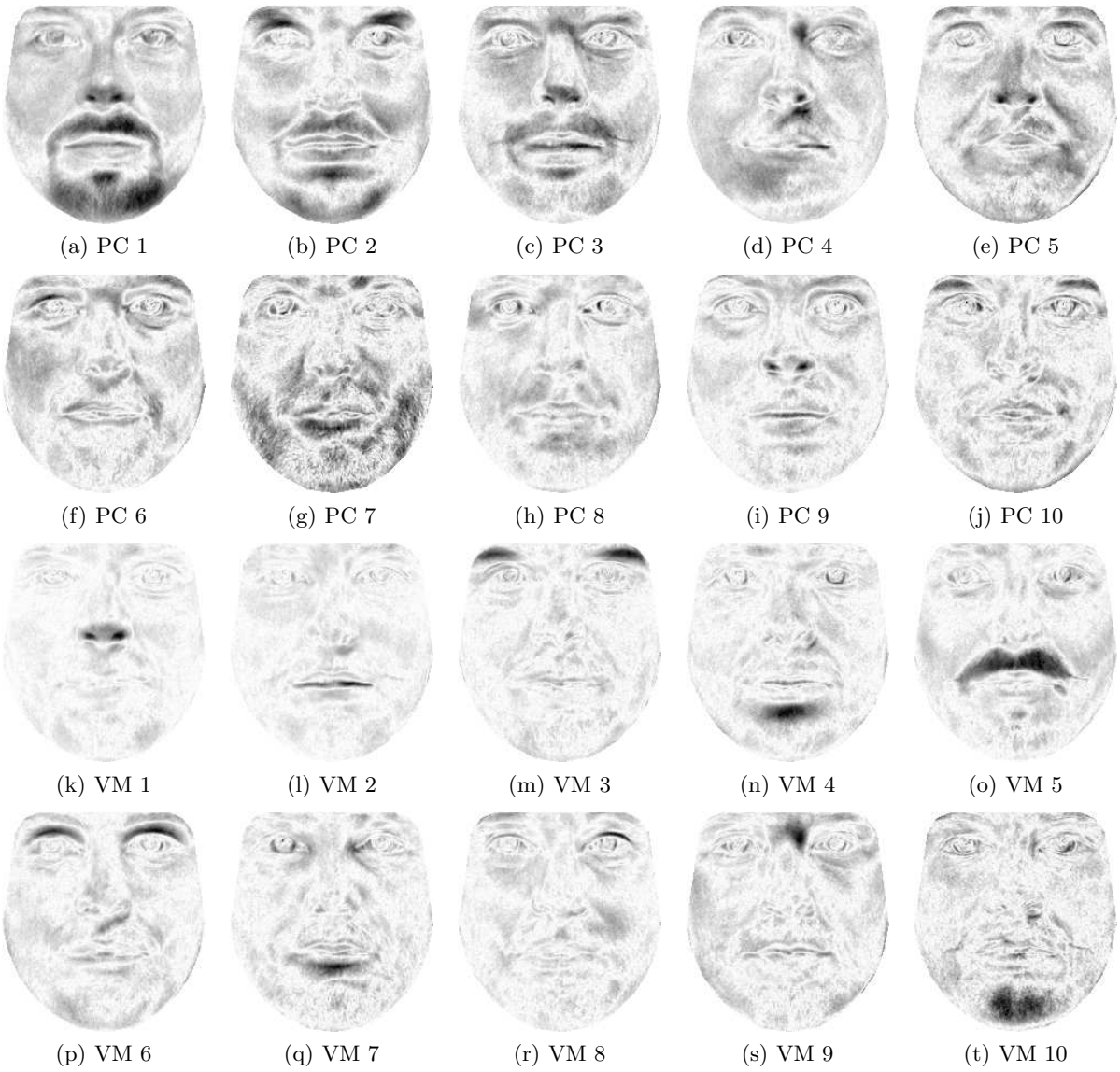

(c) Loadings (d) Mode variances (e) Correlation coefficients

**Figure 2.** Shape modes calculated from 62 corpus callosum annotations in mid-sagittal brain magnetic resonance images.

shape by a piece-wise affine image warp similar to the one usually carried out in Active Appearance Models.[44, 45] While the PCA modes in Figures 3(a–j) demonstrate several effects within each mode, the varimax modes in Figures 3(k–t) show nicely isolated effects. The first principal component for example, shows absence/presence of beard as well as nostrils. Notice that Figures 3(a–t) show the magnitude of the variable loadings of each pixel position of the model, rather than the actual values of the basis vectors. Black and white represents high and low magnitude, respectively. Varimax modes are in this case ordered according to the sparsity criterion, namely the variance of the squared loadings of a mode. Interpretations of the first five varimax modes are as follows, absence/presence of i) nostrils, ii) lip spacing, iii) eyebrow thickness/shadow, iv) shadow below lower lip, and v) mustache. Each of these modes are shown in Figure 4 as modifications of the mean texture. Here it becomes more apparent that the modes, albeit being sparse, also carries additional information outside the areas mentioned above. This further indicates that even subtle changes to the texture can carry substantial changes to the perceived identity.

Orthomax rotations have also been implemented in a complete Active Appearance Models framework[42] with the aim of assessing their potential in future registration studies. A preliminary cross-validation study in the face data set showed a slight, though presumably insignificant, increase in accuracy[†] when employing varimax rotation to the texture model, compared to standard PCA-based texture model. To this end it is important to stress that uncoupled shape and texture models must be employed. If not, the third PCA traditionally used in AAMs will diagonalize the covariance matrices and yield a combined basis identical to that of the standard AAM.

[†]Measured using the point to point distance between the ground truth shape and the converged model shape.



(a) PC 1 (b) PC 2 (c) PC 3 (d) PC 4 (e) PC 5

(f) PC 6 (g) PC 7 (h) PC 8 (i) PC 9 (j) PC 10

(k) VM 1 (l) VM 2 (m) VM 3 (n) VM 4 (o) VM 5

(p) VM 6 (q) VM 7 (r) VM 8 (s) VM 9 (t) VM 10

**Figure 3.** The magnitude of eigenvectors calculated from 37 face images arranged as eigenimages. PCA modes are ordered according to the variance of corresponding principal score. Varimax modes are ordered according to sparsity given by the variance of the squared loadings.

(a) Mean (b) VM 1 (c) VM 2 (d) VM 3 (e) VM 4 (f) VM 5

**Figure 4.** Varimax texture modes calculated from 37 face images. (a) mean texture. (b–f) mean texture modified by 2.5 standard deviations of the corresponding mode scores.

Quartimax rotations were also carried out in the two former case studies. Although the deformation modes by design should show more exclusive changes[‡] this behavior was not very clear. Due to the lack of differentiation from the varimax case, we have chosen not to show the quartimax modes.

## 5. DISCUSSION

The medical image analysis literature is surprisingly devoid of references to sparse modeling using the orthomax criterion. The main contribution of this paper is therefore three-fold, i) broadening the knowledge of this simple, yet powerful, modification of principal components, ii) discussing its merits, and iii) providing a diverse range of examples on its use in medical applications.

We have found the method to be conceptually simple to understand as well as to implement. This is partly due to being a well-understood and well-described method within factor analysis. We further note that orthomax transformations are independent of domain dimensionality (2D/3D etc.) and spatial structure. An additional benefit is that many common computational frameworks already provide an implementation, e.g. R, S-plus, Matlab, et cetera. Considering the selection of $k$ (the number of retained components) to lie with PCA, orthomax rotations are parameter-free. This is obviously a two-edged sword; while it leaves no frustrating choices up to the operator, it lacks the fine-grained flexibility, found in e.g. the sparse PCA method by Zou, Hastie and Tibshirani.[26] Compared to sparse PCA, orthomax rotations have the benefit of being computationally feasible even for very high-dimensional spaces, found in e.g. texture modeling. Unfortunately, and unlike sparse PCA, orthomax rotations will rarely provide entirely sparse components. This is also illustrated by the examples in this paper. However, the relative differences in magnitude within orthomax modes may in practice be considered sufficiently sparse in many cases. As hinted earlier, the resulting amount of sparsity is directly related to the rank of the variation and the number of principal components subjected to orthomax rotation.

A long term goal for sparse modeling in relation to image interpretation and registration is to be able to separate inherent variation sources from chance correlation, thus providing greater – and justifiable – model flexibility, and in addition provide parameterizations that capture latent structures more accurately. The latter aspect could be of crucial importance in highly flexible, non-linear regression methods sensitive to initialization.

Application-wise, we note that pathologies are typically spatially localized, either with respect to shape or texture. Thus, we anticipate many medical application areas where sparse parameterizations, similar to the presented approach, are preferable to the conventional global PCA approach.

## 6. CONCLUSION

We have explored a computationally simple approach for rotation of principal components using the orthomax criterion, which directly optimizes sparsity leading to localized modes of variation suitable for medical image interpretation and exploratory analyses. We have found that both high-dimensional sparse modeling of shape variability ($p \approx 300$), as well as extremely high-dimensional sparse modeling of texture variability ($p \approx 30000$) are feasible. Case studies on radiographs, brain MRI, and face images showed local modes of natural variation contrary to global PCA modes. Applications include computer-aided diagnosis in terms of exploratory analyses, disease characterization, et cetera.


## Acknowledgements

Dr. Bram van Ginneken, Image Sciences Institute, University Medical Center Utrecht, kindly provided the lung annotations used in this study. Dr. Ginneken also assisted with the anatomical interpretation. Charlotte Ryberg and Egill Rostrup, The Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital Hvidovre, kindly provided the MRIs used to produce the corpus callosum annotations. M. B. Stegmann was supported by The Danish Research Agency, grant no. 2059-03-0032. K. Sjöstrand was supported by The Technical University of Denmark, DTU.


[‡]In the sense that if one subpart of the shape is affected in one mode, it should not be much affected in the remaining modes.

## APPENDIX A. DETAILS ON ALGORITHM 1

This appendix serves to demonstrate the validity of Algorithm 1 in relation to Equation 1. Let $\odot$ denote the Hadamard (element-wise) product, let $\mathbf{A}_j$ denote the $j^{\text{th}}$ column of $\mathbf{A}$, and let $\mathbf{\Gamma} = \mathbf{\Lambda} \odot \mathbf{\Lambda}$ (remember $\mathbf{\Lambda} = \mathbf{\Phi}\mathbf{R}$). Further, the following two Hadamard relations[46] will be used:

Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}^\mathsf{T}$ denote $m \times n$ matrices and let $\mathbf{1}_q$ be a column vector of $q$ ones. Then

$$\text{trace}((\mathbf{A} \odot \mathbf{B})(\mathbf{C}^\mathsf{T} \odot \mathbf{D})) = \text{trace}((\mathbf{A} \odot \mathbf{B} \odot \mathbf{C})\mathbf{D}) \tag{6}$$

and

$$\mathbf{1}_m^\mathsf{T}(\mathbf{A} \odot \mathbf{B})(\mathbf{C}^\mathsf{T} \odot \mathbf{D})\mathbf{1}_m = \text{trace}(\mathbf{C}\ \text{Diag}(\mathbf{A}^\mathsf{T}\mathbf{B})\mathbf{D}). \tag{7}$$

Equation 1 can now be written in matrix form,

$$\begin{aligned}
\mathbf{R}_{orthomax} &= \arg\max_{\mathbf{R}} \left( \sum_{j=1}^{k} \sum_{i=1}^{p} \mathbf{\Gamma}_{ij}^2 - \frac{\gamma}{p} \sum_{j=1}^{k} \left( \sum_{i=1}^{p} \mathbf{\Gamma}_{ij} \right)^2 \right) \\
&= \arg\max_{\mathbf{R}} \left( \text{trace}(\mathbf{\Gamma}^\mathsf{T}\mathbf{\Gamma}) - \frac{\gamma}{p} \sum_{j=1}^{k} \left( \mathbf{1}_p^\mathsf{T}\mathbf{\Gamma}_j \right)^2 \right) \\
&= \arg\max_{\mathbf{R}} \left( \text{trace}(\mathbf{\Gamma}^\mathsf{T}\mathbf{\Gamma}) - \frac{\gamma}{p} \sum_{j=1}^{k} \mathbf{1}_p^\mathsf{T}\mathbf{\Gamma}_j\mathbf{\Gamma}_j^\mathsf{T}\mathbf{1}_p \right) \\
&= \arg\max_{\mathbf{R}} \left( \text{trace}(\mathbf{\Gamma}^\mathsf{T}\mathbf{\Gamma}) - \frac{\gamma}{p} \mathbf{1}_p^\mathsf{T}\mathbf{\Gamma}\mathbf{\Gamma}^\mathsf{T}\mathbf{1}_p \right) \\
&= \arg\max_{\mathbf{R}} \left( \text{trace}((\mathbf{\Lambda} \odot \mathbf{\Lambda})^\mathsf{T}(\mathbf{\Lambda} \odot \mathbf{\Lambda})) - \frac{\gamma}{p} \mathbf{1}_p^\mathsf{T}(\mathbf{\Lambda} \odot \mathbf{\Lambda})(\mathbf{\Lambda} \odot \mathbf{\Lambda})^\mathsf{T}\mathbf{1}_p \right) \\
&= \arg\max_{\mathbf{R}} \left( \text{trace}(\left(\mathbf{\Lambda}^\mathsf{T} \odot \mathbf{\Lambda}^\mathsf{T} \odot \mathbf{\Lambda}^\mathsf{T}\right) \mathbf{\Lambda}) - \frac{\gamma}{p}\ \text{trace}(\mathbf{\Lambda} \cdot \text{Diag}(\mathbf{\Lambda}^\mathsf{T}\mathbf{\Lambda})\mathbf{\Lambda}^\mathsf{T}) \right) \\
&= \arg\max_{\mathbf{R}} \left( \text{trace}(\mathbf{R}^\mathsf{T}\mathbf{\Phi}^\mathsf{T}(\mathbf{\Lambda} \odot \mathbf{\Lambda} \odot \mathbf{\Lambda})) - \frac{\gamma}{p}\ \text{trace}(\mathbf{\Lambda}^\mathsf{T}\mathbf{\Lambda} \cdot \text{Diag}(\mathbf{\Lambda}^\mathsf{T}\mathbf{\Lambda})) \right) \\
&= \arg\max_{\mathbf{R}} \left( \text{trace}(\mathbf{R}^\mathsf{T}\mathbf{\Phi}^\mathsf{T}(\mathbf{\Lambda} \odot \mathbf{\Lambda} \odot \mathbf{\Lambda}) - \frac{\gamma}{p}\mathbf{R}^\mathsf{T}\mathbf{\Phi}^\mathsf{T}\mathbf{\Lambda} \cdot \text{Diag}(\mathbf{\Lambda}^\mathsf{T}\mathbf{\Lambda})) \right) \\
&= \arg\max_{\mathbf{R}} \left( \ \text{trace}(\mathbf{R}^\mathsf{T}\mathbf{Q})\ \right) \quad \text{where} \quad \mathbf{Q} = \mathbf{\Phi}^\mathsf{T}(\mathbf{\Lambda} \odot \mathbf{\Lambda} \odot \mathbf{\Lambda} - \frac{\gamma}{p}\ \mathbf{\Lambda} \cdot \text{Diag}(\mathbf{\Lambda}^\mathsf{T}\mathbf{\Lambda})).
\end{aligned} \tag{8}$$

In Algorithm 1, an iterative approach is taken to solving Equation 1. Here, the part where $\mathbf{R}$ enters non-linearly, i.e. $\mathbf{Q}$, is kept fixed using the current estimate of $\mathbf{R}$. Then, the singular value decomposition, in line 6 of Algorithm 1, produces the optimal $\mathbf{R}$ for the linear part as shown in Equation 8 which subsequently replaces the current estimate. The initial estimate of $\mathbf{R}$ is the identity matrix.

By assuming that $\mathbf{Q}$ does not depend on $\mathbf{R}$, then Equation 8 would be maximized if and only if $\mathbf{R}^\mathsf{T}\mathbf{Q}$ is symmetric and positive semi-definite[§]. This can be accomplished by choosing $\mathbf{R} = \mathbf{U}\mathbf{V}^\mathsf{T}$, where $\mathbf{U}$ and $\mathbf{V}$ are taken from the singular value decomposition $\mathbf{Q} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$. That $\mathbf{R}^\mathsf{T}\mathbf{Q}$ is symmetric and positive semi-definite can been seen by the following substitution[¶]:

$$\mathbf{R}^\mathsf{T}\mathbf{Q} = \mathbf{R}^\mathsf{T}\mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T} = (\mathbf{U}\mathbf{V}^\mathsf{T})^\mathsf{T}\mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T} = \mathbf{V}\mathbf{U}^\mathsf{T}\mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T} = \mathbf{V}\mathbf{S}\mathbf{V}^\mathsf{T}. \tag{9}$$

This concludes our presentation of the background of Algorithm 1 based on Refs. 46 and 47. Further details can be found in Refs. 33–35, 46 and 48.

[§] See the compact proof of Theorem 2 in Ref. 47 on Procrustes analysis.

[¶] Remember that $\mathbf{S}$ is a diagonal matrix of singular values, and $\mathbf{U}$ and $\mathbf{V}$ hold orthogonal singular vectors.

## REFERENCES


1. S. C. Mitchell, B. P. F. Lelieveldt, H. G. Bosch, J. H. C. Reiber, and M. Sonka, "Disease characterization of active appearance model coefficients," *Proc. of SPIE - The International Society for Optical Engineering* **5032 II**, pp. 949–957, 2003.
2. M. Üzümcü, A. F. Frangi, J. H. C. Reiber, and B. P. F. Lelieveldt, "Independent component analysis in statistical shape models," in *Proc. of SPIE - The International Society for Optical Engineering*, **5032 I**, pp. 375–383, 2003.
3. M. Üzümcü, A. F. Frangi, M. Sonka, J. H. C. Reiber, and B. P. F. Lelieveldt, "ICA vs. PCA active appearance models: Application to cardiac MR segmentation," in *MICCAI 2003, Montréal, Canada*, *LNCS* **2878**, pp. 451–458, 2003.
4. A. Suinesiaputra, M. Üzümcü, A. F. Frangi, T. A. M. Kaandorp, J. H. C. Reiber, and B. P. F. Lelieveldt, "Detecting regional abnormal cardiac contraction in short-axis MR images using independent component analysis," in *Lecture Notes in Computer Science*, **3216**, pp. 737–744, Springer-Verlag, 2004.
5. A. Suinesiaputra, A. F. Frangi, M. Üzümcü, J. H. C. Reiber, and B. P. F. Lelieveldt, "Extraction of myocardial contractility patterns from short-axes MR images using independent component analysis," in *Lecture Notes in Computer Science*, **3117**, pp. 75–86, Springer-Verlag, 2004.
6. P. D. Sozou, T. F. Cootes, C. J. Taylor, and E. C. Di Mauro, "Non-linear generalization of point distribution models using polynomial regression," *Image and Vision Computing* **13**(5), pp. 451–7, 1995.
7. T. F. Cootes and C. J. Taylor, "A mixture model for representing shape variation," in *Proceedings of the 8th British Machine Vision Conference, BMVC*, **1**, pp. 110–119, BMVA Press, 1997.
8. T. F. Cootes and C. J. Taylor, "A mixture model for representing shape variation," *Image and Vision Computing* **17**(8), pp. 567–574, 1999.
9. T. Heap and D. Hogg, "Improving specificity in PDMs using a hierarchical approach," 1997.
10. R. Bowden, *Learning non-linear Models of Shape and Motion.* PhD thesis, Dept Systems Engineering, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK., 1999.
11. S. Romdhani, S. Gong, and A. Psarrou, "A multi-view nonlinear active shape model using kernel PCA," *BMVC99. Proceedings of the 10th British Machine Vision Conference* , pp. 483–92 vol.2, 1999.
12. C. J. Twining and C. J. Taylor, "Kernel principal component analysis and the construction of non-linear active shape models," in *Proceedings of the British Machine Vision Conference, BMVC*, **1**, pp. 23–32, 2001.
13. R. Larsen, "Shape modelling using maximum autocorrelation factors," in *Proceedings of the Scandinavian Image Analysis Conference (SCIA'01)*, I. Austvoll, ed., pp. 98–103, (Bergen, Norway), jun 2001.
14. R. Larsen, "Decomposition using maximum autocorrelation factors," *Journal of Chemometrics* **16**(8-10), pp. 427–435, 2002.
15. R. Larsen, H. Eiriksson, and M. B. Stegmann, "Q-MAF shape decomposition," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001, 4th International Conference, Utrecht, The Netherlands*, W. J. Niessen and M. A. Viergever, eds., *LNCS* **2208**, pp. 837–844, Springer, 2001.
16. P. Switzer, "Min/max autocorrelation factors for multivariate spatial imagery," in *Computer Science and Statistics*, L. Billard, ed., pp. 13–16, Elsevier Science Publishers B.V., 1985.
17. K. B. Hilger, M. B. Stegmann, and R. Larsen, "A noise robust statistical model for image representation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2002, 5th Int. Conference, Tokyo, Japan*, *LNCS* **2488**, pp. 444–451, sep 2002.
18. K. B. Hilger, R. Larsen, and M. Wrobel, "Growth modeling of human mandibles using non-Euclidean metrics," *Medical Image Analysis* **7**, pp. 425–433, 2003.
19. R. Larsen and K. B. Hilger, "Statistical 2D and 3D shape analysis using non-Euclidean metrics," *Medical Image Analysis* **7**(4), pp. 417–423, 2003.
20. R. Larsen, K. B. Hilger, and M. C. Wrobel, "Statistical 2D and 3D shape analysis using non-Euclidean metrics," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002, 5th Int. Conference, Tokyo, Japan*, *Lecture Notes in Computer Science* **2489**, pp. 428–435, Springer, 2002.
21. L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters* **72**(23), pp. 3634–3637, 1994.
22. J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer Verlag, 1997.
23. B. S. Peterson, P. A. Feineigle, L. H. Staib, and J. C. Gore, "Automated measurement of latent morphological features in the human corpus callosum," *Human Brain Mapping* **12**(4), pp. 232–245, 2001.
24. C. Chennubhotla and A. Jepson, "Sparse PCA extracting multi-scale structure from data," *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on* **1**, pp. 641–647, 2001.
25. H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psychometrika* **23**, pp. 187–200, 1958.
26. H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," tech. rep., Statistics Department, Stanford University, 2004.
27. H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2), pp. 301–320, 2005.

28. I. T. Jolliffe, N. T. Trendafilov, and M. Uddin, "A modified principal component technique based on the LASSO," *Journal of Computational and Graphical Statistics* **12**(3), pp. 531–547, 2003.
29. K. Sjöstrand, M. B. Stegmann, and R. Larsen, "Sparse principal component analysis in medical shape modeling," in *International Symposium on Medical Imaging 2005, San Diego, CA, USA*, **6144**, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, (Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby), feb 2006.
30. T. F. Cootes, D. Cooper, C. J. Taylor, and J. Graham, "A trainable method of parametric shape description," in *Proc. British Machine Vision Conference*, pp. 54–61, Springer-Verlag, 1991.
31. T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham, "Training models of shape from sets of examples," in *Proc. British Machine Vision Conf., BMVC92*, pp. 9–18, 1992.
32. H. H. Harman, *Modern Factor Analysis*, The University of Chicago Press, 1967.
33. P. Horst, *Factor Analysis of Data Matrices*, Holt, Rinehart and Winston, New York, 1965.
34. J. M. F. ten Berge, "A joint treatment of varimax rotation and the problem of diagonalizing symmetric matrices simultaneously in the least-squares sense," *Psychometrika* **49**(3), pp. 347–358, 1984.
35. R. J. Sherin, "A matrix formulation of Kaiser's varimax criterion," *Psychometrika* **31**(4), pp. 535–538, 1966.
36. I. T. Jolliffe, "Rotation of principal components: Choice of normalization constraints," *Journal of Applied Statistics* **22**(1), pp. 29–36, 1995.
37. I. T. Jolliffe, "Rotation of ill-defined principal components," *Applied Statistics* **38**(1), pp. 139–147, 1989.
38. M. B. Stegmann, *Generative Interpretation of Medical Images*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, 2004.
39. B. van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database," *Medical Image Analysis* , 2005.
40. M. B. Stegmann, K. Skoglund, and C. Ryberg, "Mid-sagittal plane and mid-sagittal surface optimization in brain MRI using a local symmetry measure," in *International Symposium on Medical Imaging 2005, San Diego, CA, Proc. of SPIE vol. 5747*, SPIE, feb 2005.
41. M. B. Stegmann, R. H. Davies, and C. Ryberg, "Corpus callosum analysis using MDL-based sequential models of shape and appearance," in *International Symposium on Medical Imaging 2004, San Diego CA, SPIE*, pp. 612–619, SPIE, feb 2004.
42. M. B. Stegmann, B. K. Ersbøll, and R. Larsen, "FAME – a flexible appearance modelling environment," *IEEE Trans. on Medical Imaging* **22**(10), pp. 1319–1331, 2003.
43. M. B. Stegmann, "Analysis and segmentation of face images using point annotations and linear subspace techniques," Tech. Rep. IMM-REP-2002-22, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, aug 2002.
44. T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proc. of European Conf. on Computer Vision 1998*, *Lecture Notes in Computer Science* **1407**, pp. 484–498, Springer, 1998.
45. T. F. Cootes and C. J. Taylor, *Statistical Models of Appearance for Computer Vision*, Tech. report, University of Manchester, 2001.
46. H. Neudecker, "On the matric formulation of kaiser's varimax criterion," *Psychometrika* **46**(3), pp. 343–345, 1981.
47. J. M. F. ten Berge, "Orthogonal procrustes rotation for two or more matrices," *Psychometrika* **42**, pp. 267–276, 1977.
48. J. M. F. ten Berge, D. L. Knol, and H. A. L. Kiers, "A treatment of the orthomax rotation family in terms of diagonalization, and a re-examination of a singular value approach to varimax rotation," *Computational Statistics Quarterly* **3**, pp. 207–217, 1988.