

Computing Minimal Deformations: Application to Construction of Statistical Shape Models

Supplementary Material

Darko Zikic¹, Michael Sass Hansen², Ben Glocker¹,
Ali Khamene³, Rasmus Larsen², and Nassir Navab¹

¹ Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany
{zikic, glocker, navab}@in.tum.de

² Informatics and Mathematical Modeling (IMM), Technical University of Denmark
{msh, rl}@imm.dtu.dk

³ Imaging and Visualization Department, Siemens Corporate Research (SCR), Princeton, USA
ali.khamene@siemens.com

A. Details of the Umeyama Method

The absolute orientation method of Umeyama [2] computes the similarity transformation which minimizes the mean squared distance between two point sets A and B of arbitrary dimension d , that is

$$e^2(R, t, c) = \frac{1}{n} \sum_{i=1}^n \|B_i - (cRA_i + t)\|^2, \quad (1)$$

where c is the scaling factor, t is the translation vector, and R is the rotation matrix, and n is the number of points.

The following outline follows closely the excellent description given in [2]. The single entities are computed based on the SVD decomposition of the covariance matrix

$$C_{AB} = \frac{1}{n} \sum_{i=1}^n (B_i - \mu_B)(A_i - \mu_A)^\top \quad (2)$$

of the n corresponding points, such that $C_{AB} = UDV^\top$. Here, μ_A and μ_B are the respective mean values and the ascending non-negative values of the diagonal matrix D are denoted by d_i . The standard deviations are given by σ_A and σ_B .

The matrix S is defined in the following way.

$$S = \begin{cases} I & \det(C_{AB}) \geq 0 \\ \text{diag}(1, 1, \dots, -1) & \det(C_{AB}) < 0 \end{cases}, \quad (3)$$

where I denotes the identity matrix.

When $\text{rank}(C_{AB}) \geq d - 1$, the similarity parameters are given by

$$R = USV^\top \quad (4)$$

$$t = \mu_B - cR\mu_A \quad (5)$$

$$c = \frac{1}{\sigma_A^2} \text{tr}(DS). \quad (6)$$

In the Equation (4), the matrix S is defined depending on the rank of the covariance matrix C_{AB} .

If $\text{rank}(C_{AB}) = d - 1$ then, S is defined as

$$S = \begin{cases} I & \det(U) \det(V) = 1 \\ \text{diag}(1, 1, \dots, -1) & \det(U) \det(V) = -1 \end{cases}. \quad (7)$$

Otherwise, the definition from (3) is used.

For the proofs of the optimality of the above estimation, please refer to the original paper [2].

B. Discussion of PCA on Limited Number of Samples, and Consequences for Non-Minimal Deformations

In the following we discuss some details of the general behavior of the PCA for finite number of samples B.1. We see that for a limited number of samples, not the actual modes generating the data, but their linear combinations are reconstructed. From this general observation, we proceed in Section B.2 by applying this to the case of non-minimal deformations, which can be seen as composed of linear similarity transformations and nonlinear deformation

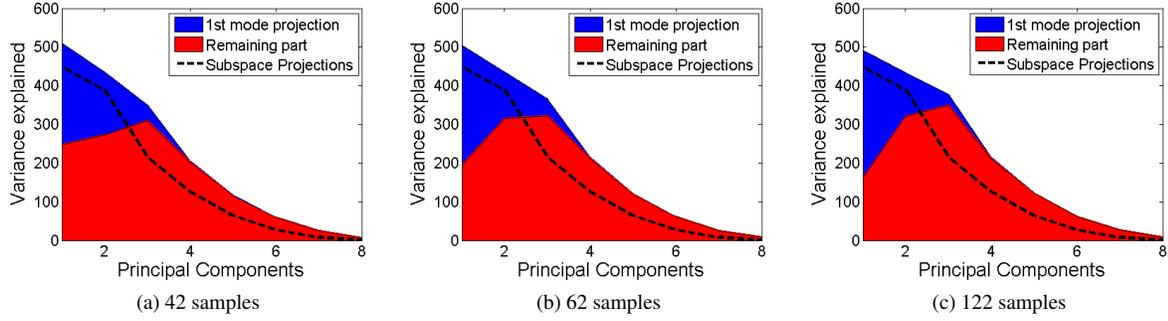


Figure 1: Example showing how the variance of first linear mode (blue area) is mixed with the remaining variance (red area) in the first modes, if a finite number of samples is used. The black curve shows the true variance of the 2nd to last modes.

modes. We conclude that in general it can be expected, that the similarity transformation is mixed with the first nonlinear modes of the resulting model, thus corrupting the ability of the model to describe shape.

B.1. PCA Based on a Limited Number of Samples

An alternative formulation of PCA than the minimization of reconstruction error, is to search for the orthonormal projections that maximize the variance of data projected onto the subspace spanned by the eigenvectors.

$$\max_{\mathbf{w}} E\{(\mathbf{w}^\top \mathbf{u})^2\} , \text{ where } \|\mathbf{w}\| = 1 , \quad (8)$$

where the components \mathbf{w} are estimated sequentially, with the constraint that they should be orthogonal to each other.

Consider the true deformation field \mathbf{u}_* to be a stochastic variable formed by a linear combination of uncorrelated factors (since the shape modeling is independent of the mean field, we will for simplicity of notation assume that the mean field is zero in the rest of the derivation).

$$\mathbf{u}_* = \mathbf{V}_* \Lambda_*^{1/2} \mathbf{b}_* , \quad (9)$$

where \mathbf{V}_* consists of a set of orthonormal vectors, the so-called modes, $\Lambda_*^{1/2}$ is a diagonal matrix with diagonal elements $\lambda_{*i}^{1/2}$, which in effect scale the components, and \mathbf{b}_* denotes a stochastic vector with uncorrelated elements of unit variance. We remark that this representation of the mixture can be achieved for all linear combination models, and as a consequence adds no further constraints.

When we estimate the modes, it is the optimum projection \mathbf{w} that we are looking for. When applying the described assumptions on \mathbf{u} we get

$$E\{(\mathbf{w}^\top \mathbf{u})^2\} = E\{(\mathbf{w}^\top \mathbf{V}_* \Lambda_*^{1/2} \mathbf{b}_*)^2\} \quad (10)$$

$$= \mathbf{w}^\top \mathbf{V}_* \Lambda_*^{1/2} E\{\mathbf{b}_* \mathbf{b}_*^\top\} \Lambda_*^{1/2} \mathbf{V}_*^\top \mathbf{w} \quad (11)$$

and since \mathbf{b}_* is a vector of uncorrelated variables, in theory $E\{\mathbf{b}_* \mathbf{b}_*^\top\} = I_p$, where I_p is the identity matrix with dimension p , and the optimization problem becomes

$$\max_{\mathbf{w}} \mathbf{w}^\top \mathbf{V}_* \Lambda_* \mathbf{V}_*^\top \mathbf{w} , \quad (12)$$

We see that the PCA solution to (12) is actually the generating model (9) itself, where $[\mathbf{w}_1, \dots, \mathbf{w}_n] = \mathbf{V}$.

However, in real applications the expectancy of the covariance matrix $\mathbf{b}_* \mathbf{b}_*^\top$ would be calculated by

$$E\{\mathbf{b}_* \mathbf{b}_*^\top\} \approx \frac{1}{N-1} \sum_{i=1}^N \mathbf{b}_i \mathbf{b}_i^\top = \mathbf{C}_b \neq I_p . \quad (13)$$

which is the unbiased estimator under assumption of normal distributed vectors. It can be seen that in spite of uncorrelated elements of the vector \mathbf{b}_* the sum in (13) will not be an identity matrix. Soper described in 1913 the distribution of correlation coefficients, and a not so surprising conclusion that can be drawn from that is, that though the correlation between two uncorrelated variables is zero in average, it is more likely, in the case of a limited number of samples, that it actually attains a value different from zero [1]. Regarding the distribution of correlation coefficients, it can be stated that for actual measurements, though the correlation between two uncorrelated variables is zero in average, it is more likely - in the case of a limited number of samples - that it actually attains a value different from zero. Let \mathbf{w}_V be the projection of the estimated mode on to the true modes \mathbf{V}_* of the data, then the estimation of the principal modes can be written as

$$\max_{\mathbf{w}_V} \mathbf{w}_V^\top \Lambda_*^{1/2} \mathbf{C}_b \Lambda_*^{1/2} \mathbf{w}_V , \quad (14)$$

From this we observe that, in general, \mathbf{w} is not just projected onto *one* component (meaning only one element of \mathbf{w}_V would be 1 and the rest 0), but \mathbf{w}_V is rather found as a

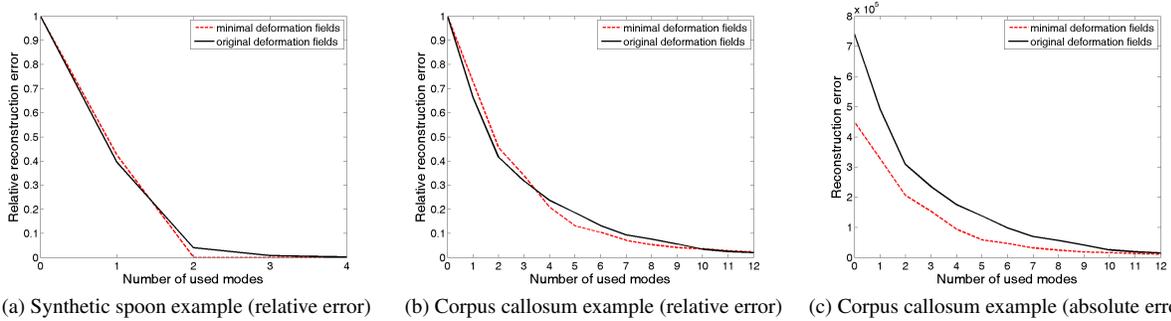


Figure 2: Analysis of the *relative* reconstruction error of the original and modified shape model on (a) synthetic and (b) on the corpus callosum data (NRM 1). Please note that the error is scaled by the initial error, such that both plots start with unit error. This is done to have a fair comparison, since the absolute error for the original model is always larger, compare (c). Notice the better reconstruction ability of the modified model when more than the very first modes are used.

mixture of all modes, for instance calculated as the eigenvectors of the $\Lambda_*^{1/2} \mathbf{C}_b \Lambda_*^{1/2}$ matrix. Still, though, the estimated correlation coefficients will tend to be small, and the first mode will normally have a larger component of the biggest variance true mode than the others. An analytical analysis of distributions of projections and blending of modes is beyond the scope of the current paper, but in Figure 1 the distribution of the estimated PCA projection on the first and largest modes of the data model is illustrated for different numbers of samples.

B.2. The Estimated Impact of Similarity Transforms in the PCA Model

A similarity transformation \mathbf{u}_l can also be described by a sum of linear deformation components. These can also without adding constraints be represented similarly to (9)

$$\mathbf{u}_l = \mathbf{V}_l \Lambda_l^{1/2} \mathbf{b}_l, \quad (15)$$

with parallel definitions of the quantities, only the subindex l indicates that the model describes only similarity transforms. Now our observed deformation field, \mathbf{u} , resulting from the nonlinear registration, can be modeled in approximation as a sum of the true deformation fields \mathbf{u}_* and an additional global similarity transform field \mathbf{u}_l as

$$\mathbf{u} = \mathbf{u}_l + \mathbf{u}_*. \quad (16)$$

We use the models of each composed field and as an expression for our observed field we get

$$\mathbf{u} = \mathbf{V}_l \Lambda_l^{1/2} \mathbf{b}_l + \mathbf{V}_* \Lambda_*^{1/2} \mathbf{b}_* = \mathbf{V} \Lambda^{1/2} \mathbf{b}, \quad (17)$$

where \mathbf{V} , Λ and \mathbf{b} denote the concatenation through matrix operations of the sum into a combined model with the same structure as the true deformation model and the similarity transform deformation model. Now the important thing to

note is that since \mathbf{u}_l is *only* describing global deformation changes, it is seen that all modes in Φ_l are also orthogonal to all modes in Φ_* (the nonlinear modes *without* similarity transforms).

In summary this means that all modes in V are therefore orthogonal, and obviously \mathbf{b}_* and \mathbf{b}_l are also uncorrelated and as a consequence the triplet \mathbf{V} , Λ and \mathbf{b} satisfy all assumptions from (9) and all the observations and conclusions from Section B.1 also apply for this model. Now consider that similarity transforms are global and thus easily describe a big amount of variance, then we may expect that the biggest mode in this model (in terms of explained variance) is indeed a similarity transform mode, and we can then expect that Figure 1 actually describes the distribution of the biggest similarity transform mode - if we do not extract it before hand.

To conclude, we have learned that similarity transform and nonlinear modes will be blended when only a finite number of samples are available, and if the similarity transform is significant we expect it to be distributed as depicted in Figure 1.

C. Reconstruction Ability of the Model

In this section we show that not only the variance, but also the relative reconstruction error of the modified SSMs is lower.

The shape models built from minimal deformations have a reduced variance compared to the original models - this is equivalent to a larger reconstruction error, and shown in Fig. 2c. We also investigate how much of the *remaining* variance can be explained by using the model and to this end we compute the *relative* reconstruction error. Here, the reconstruction error is divided by the norm of original error, in order to enable the comparison of relative performance of original and modified model. Please note that without this

scaling, the original model would yield significantly larger errors than the modified version.

The reconstruction ability is tested by averaging a series of 10 leave-10-out experiments. This is done for the synthetic spoon example as well as for the corpus callosum data set. It can be seen in Fig. 2 that the very first modes of the original model describe more than the corresponding modes of the modified model. When more than the very first modes are included - which is the interesting case for applications - the reduction of the relative error of the modified model is superior, compare Fig. 2.

Please also note that for the simple synthetic example which was generated from two modes, the relative reconstruction error is reduced to zero by the modified model by using the first two modes, in contrast to the original model, compare Fig. 2a.

References

- [1] H. E. Soper. On the probable error of the correlation coefficient to a second approximation. *Biometrika*, 9:91–115, 1913. 2
- [2] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(4):376–380, 1991. 1