Optimal Sub-Shape Models by Minimum Description Length

Paper Number : 1060

Abstract

Active shape models are a powerful and widely used tool to interpret complex image data. By building models of shape variation they enable search algorithms to use a priori knowledge in an efficient and gainful way. However, due to the linearity of PCA, non-linearities like rotations or independently moving subparts in the data can deteriorate the resulting model considerably.

Although non-linear extensions of active shape models have been proposed and application specific solutions have been used, they still need a certain amount of user interaction during model building. In this paper the task of building/choosing optimal models is tackled in a more generic information theoretic fashion. In particular, we propose an algorithm based on the minimum description length principle to find an optimal subdivision of the data into subparts, each adequate for linear modeling. This results in an over all more compact model configuration. Which in turn leads to a better model in terms of modes of variations. The proposed method is evaluated on synthetic data, medical images and hand contours.

1 Introduction

Introduced in 1992 *active shape models (ASMs)* [1] have proven to be very successful in interpreting complex image data. Due to noise, overlapping structures of varying shape, and the need to consistently identify instances of anatomical structures in a large number of images of potentially different patients, the use of a priori knowledge is particularly useful in medical imaging [2]. Furthermore it can be used to adopt a notion of healthy versus pathologically altered shapes [3]. ASMs use PCA to build a linear model of shape variation from a set of training examples.

Since non-linearities violate the linearity assumption of PCA, they degrade the compactness and thus the efficiency of the model. Non-linearities can occur among other reasons due to rotation of distinct anatomical structures. Various approaches to deal with specific non-linearities have been proposed. In [4, 5] they were dealt with by polynomial regression or multi-layer perceptrons. However the order of polynomials or the architecture of the network had to be chosen application specific. In [6] mixture models were used resulting in more reliable models but becoming un-feasible for large training set sizes. In [3] snakes were used to deal with pathological local non-linearities during the search procedure.

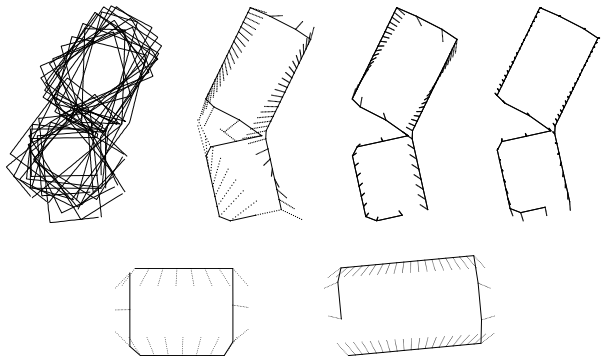

Figure 1: Upper row: Aligned set of rotating rectangles, each with changing aspect ratio, and first 3 modes of variation. Lower row: first modes for the separate rectangles (the small lines indicate the modes of variations for each landmark point).

In Fig. 1 a simple example of non-linear shape variation is depicted. Two rectangles rotate against each other while independently changing the aspect ratio. In the upper row the aligned shape set and the first 3 modes of variation resulting from the entire shape are depicted. The modes are visualized by the mean shape and lines indicating the deformation caused by the modes of shape variation. Note that aspect ratio and rotation changes interact with each other and deteriorate the compactness of the model considerably. In the lower row the two rectangles are modeled separately and the change of aspect ratio is plausibly represented in the first modes. As an alternative to modeling the non-linear variations a correct determination of distinct entities in the

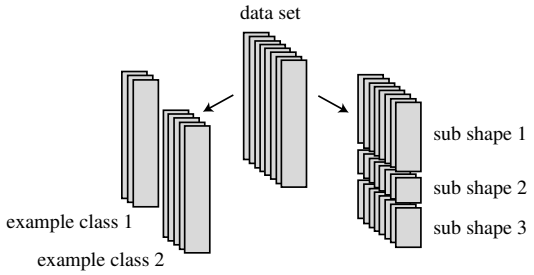

Figure 2: Data vectors in the training set, when the training set is split into subsets one obtains a mixture model (left part of figure); when the vectors are split one obtains the model presented in this paper (right part of figure).

data seems to be a worthwhile and crucial step to build compact and efficient models.

Minimum description length (MDL), a model selection criterion has been used successfully in different applications. It gives a criterion to compare likelihoods of models that describe a given data.

In [7, 8] MDL was used to establish landmark correspondences on a set of shapes defined by continuous contours before active shape model training was performed. In [9] MDL was used for group-wise non-rigid registration. A method proposed in [10] uses MDL to select hypotheses for robust appearance based object recognition. In [11] multiple eigenspaces were build in order to account for groups of different objects present in a training set, thereby improving recognition results, using better and more specialized models.

The main contribution of this work is a novel method for obtaining compact shape models. In particular we propose an efficient MDL based method that identifies in a fully automatic manner the number of sub-shapes and their locations in order to represent a given set of training shapes compactly. As the experiments will show the method is able to identify the distinct sub parts of a shape automatically. Overall this results in more compact models which in turn lead to a better generalization behavior.

This attempt goes along with recent developments in model based approaches that aim at more autonomous model generation processes with decreasing amount of user interaction. In a sense our concept is complementary to the approach examined in [6, 11], where as illustrated in Fig. 2 not the data vectors but the data set is split.

The application of the method is the more and more unsupervised model generation, where algorithms have to be able to identify distinct entities in order to make efficient modeling possible. The proposed method is able to identify distinct sub parts of shapes automatically (e.g., the fingers of a hand). Therefore, the tedious, time consuming and sometimes even sub optimal process of manual splitting can be avoided.

This paper is structured as follows: in section 2 a criterion based on MDL for the selection of sub-shape models will be derived, and a search procedure using this criterion to find an optimal set of models will be explained. Section 3 demonstrates results of the algorithm on artificial and real data and section 4 concludes with a discussion.

2 A criterion for multiple shape model selection

In order to find an optimal model ensemble describing the data set the minimum description length principle will be used [12, 13]. It states that maximizing the likelihood of a model $\mathcal{M}$ given certain data D is equivalent to minimizing the cost of communicating the model itself and the data encoded with help of the model i.e.

$$L(D, \mathcal{M}) = L(\mathcal{M}) + L(D|\mathcal{M}). \tag{1}$$

It is a computationally more feasible formulation of Occams razor or assuming that *shortest explanations generalize best.*

We will use ASMs for modeling the shape data. Subsequently an MDL criterion will be used to judge the encoding of the shape data with ensembles of sub-shape ASMs, to ultimately obtain an optimal sub-division of the data.

Statistical shape models The principle ASM concept is based on shapes that are represented by a finite set of n landmarks. Each shape in the training set can be represented by a $2n$ dimensional vector $\mathbf{x}_i$ generated by concatenation of the x and y coordinates in 2 dimensional data (extensions to 3D are straightforward). In order to achieve a compact representation PCA is used on the set $\{\mathbf{x}_i, i = 1, \ldots, n_T\}$ and thereby a new coordinate system that represents each of the vectors

$$\mathbf{x}_i = \bar{\mathbf{x}} + \sum_{j=1}^{n_p} a_j \mathbf{e}_j, \tag{2}$$

in an optimal way is created. The modes $\mathbf{e}_j$ are the eigenvectors of the covariance matrix sorted according to decreasing eigenvalue λ_j. $\bar{\mathbf{x}}$ is the mean shape and n_p can be chosen to fulfill a given accuracy constraint. The eigenvalues λ_j correspond to the variance of the data in the direction $\mathbf{e}_j$.

2.1 Description length of shape models

If modeling shape data by a multivariate Gaussian in the directions of the eigenvectors $\mathbf{e}_j$ as described above we can apply Shannons theorem [14] to each of these 1D distributions. The corresponding coefficients a_j^i are quantized

by the step size Δ_{Im} which is related to the pixel-size, and are strictly bounded by R_j. For each training sample $\mathbf{x}_i$ the new discrete coordinates $\hat{a}_j^i = k\Delta_{Im}, k \in \mathbb{Z}$ with $-R_j/2 \leq \hat{a}_j \leq R_j/2$ are modeled by a Gaussian distribution with coefficient mean value $\mu_j = 0$ and standard deviation $\sigma_j = \sqrt{\lambda_j}$.

For each dimension j of the eigenspace used to encode the data the transmission costs of the model $L(\mathcal{M}_{\mathbf{e_j}})$ are the quantized eigenvector, $\hat{\sigma_j}$ and the quantization parameter δ_j for the direction $\mathbf{e}_j$. $L(D|\mathcal{M}_{\mathbf{e_j}})$ is the cost of transmitting the data i.e. the quantized coefficients $\hat{a}_j^i$ of the training set with respect to the direction $\mathbf{e}_j$.

The description length for the data encoded with an n_p dimensional eigenspace is the sum of the transmission costs for the data encoded using the eigenvectors $(\mathbf{e}_j)_{j=1,\ldots,n_p}$ together with the cost of the residual error

$$\sum_{j=1}^{n_p} \left(L(\mathcal{M}_{\mathbf{e_j}}) + L(D|\mathcal{M}_{\mathbf{e_j}}) \right) + \mathcal{R}. \tag{3}$$

For a detailed derivation of the description length of 1D Gaussians please refer to [7].

Multiple models A set of models $\{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$ each representing a part of the data D with every part of the data covered by at least one model will be called a *model ensemble* $\mathsf{M} = \langle \mathcal{M}_1, \ldots, \mathcal{M}_n; \mathcal{S} \rangle$, where $\mathcal{S}$ holds the information of the data parts corresponding to the individual sub-models. The minimization of

$$C(\mathsf{M}) = L(\mathcal{S}) + \sum_{\mathcal{M}_i \in \mathsf{M}} L(\mathcal{M}_i) + L(D_i|\mathcal{M}_i) + \mathcal{R}, \tag{4}$$

where $L(\mathcal{S})$ is the additional cost for transmitting the splitting information, then corresponds to the maximization of the likelihood of the model ensemble.

2.2 The Criterion function

After establishing the general criterion this section will concentrate on the application to eigenspace models. First we will derive the criterion function which will allow for a search through the possible sub-shape model ensembles aiming at an optimal data partition in the sense of compactness of the generated over all model. While the primary thread will refer to shape models the application to any vector data is straight forward. When Eq. 4 is applied to an ensemble M of eigenspace models $\mathcal{M}_m$ the criterion becomes

$$C(\mathsf{M}) = L(\mathcal{S}) + \sum_{m:\mathcal{M}_m \in \mathsf{M}} \left(\sum_{j=1}^{n_p^m} \mathcal{C}_j^m + \mathcal{R} \right), \tag{5}$$

where $n_p^m = \max\{j : \sigma_j > \Delta_{Im}\}$ is the dimension of the utilized eigenspace. $L(\mathcal{S})$ is the additional cost to transmit the split information. This term acts as a penalty for additional splits and prohibits possible trivial solutions. In the case of shapes $L(\mathcal{S})$ is the cost of identifying the indices of n landmarks where a split takes place. Assuming $l \geq 2$ sub-models and equal probability for all possible split positions, the cost is

$$L(\mathcal{S}) = l \cdot \log_2(n) \tag{6}$$

$\mathcal{C}_j^m$ is the coding cost term for the jth eigendirection of the eigenspace $\mathcal{M}_m$

$$\begin{aligned} \mathcal{C}_j^m &= 1 + \log_2\left(\frac{\sigma_{max} - \sigma_{min}}{\delta_j}\right) + |\log_2 \delta_j| - & (7) \\ &\quad -n_T \log_2 \Delta_{Im} + \frac{n_T}{2} \log_2(2\pi\sigma_j^2) + \\ &\quad + \frac{n_T}{2} + \frac{n_T \delta_j^2}{12\sigma_j^2}, & (8) \end{aligned}$$

where $\sigma_{max} = R/2$ and $\sigma_{max} = 2\Delta_{Im}$. $\mathcal{R}$ is the residual error that remains after fitting the training set with the model

$$\mathcal{R} = n_T \sum_{j=n_p+1}^{2n} \lambda_j. \tag{9}$$

In order to express preferences (e.g., more sub-shape models) one can introduce a weighting parameter k which decreases/increases the splitting costs.

2.3 Search for an optimal set of sub shape models

The criterion function $C(\mathsf{M})$ in Eq. 5 has to be integrated into a search procedure in order to determine an optimal splitting of the shape data. In Fig. 3 the value of the criterion function is depicted relative to its outcome for a single model for all possible splits of a set of bone contours into two sub-shapes. For better visibility only 32 landmarks were used. x and y coordinates indicate the indices of landmarks where the shape is split. The global minimum lies at the rotation point between the two bones: $C(\langle \mathcal{M}_1, \mathcal{M}_2, \mathcal{S} \rangle)/C(\mathcal{M}_{total}) = 0.841$.

Nevertheless a global search is not feasible for larger shape vectors and the possibility of multiple splitting positions. For our experiments we employed a hypothesize and select procedure.

1. The search is initiated with small *atoms* i.e. sub-vectors of the data vectors $\mathbf{x}_i$. For shapes landmarks can be used, for images an initial segmentation can be applied to the image data in order to decrease computation costs. Neighborhood relations are established accordingly.

2. In a local growing process the sub-shapes merge guided by $C(\mathsf{M})$: A randomly chosen sub-shape model $\mathcal{M}_i$ merges with its neighbor $\mathcal{M}_j$ if for the new ensemble $\mathsf{M}' =$

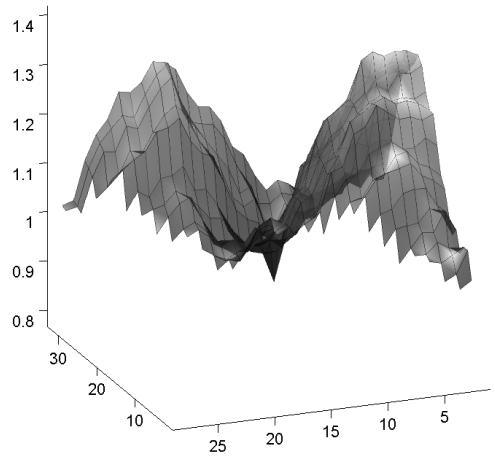

Figure 3: $C(\mathsf{M})/C(\mathcal{M}_{total})$ for bone contours as function of split positions if split into two sub-shapes. Note the deep minimum (0.841) at the correct split position.

$\langle \ldots, \mathcal{M}_{i\cup j}, \ldots, \mathcal{S}' \rangle$, $C(\mathsf{M}')$ is the minimum for all neighbors of $\mathcal{M}_i$ and $C(\mathsf{M}') < C(\mathsf{M})$ holds, where $\mathsf{M} = \langle \ldots, \mathcal{M}_i, \mathcal{M}_j, \ldots, \mathcal{S} \rangle$ i.e. the criterion value decreases. For this comparison $C(\mathsf{M}')$ only has to be recalculated locally since the description length of the sub-shape models and the splitting costs contribute additively to the criterion function.

3. The n sub-shape ensembles M_k resulting from the growing processes are treated as hypotheses that are subject to a final selection procedure yielding the final result $\mathsf{M}_{\mathsf{sel}}$: $C(\mathsf{M}_{\mathsf{sel}}) = \min\{C(\mathsf{M}_k), k = 1, \ldots, n\}$.

3 Experimental Results

Experiments were performed on three different data sets: synthetic data consisting of two rectangles independently changing aspect ratios and rotating against each other, bone contours of the *metacarpal 5* and *proximal phalange 5* and on hand outlines. With $\Delta_{im} = 1$ the hypothesize-select search was performed and the resulting sub-shape models were evaluated with respect to their compactness and the resulting reconstruction error.

Synthetic data The synthetic data already referred to in the introduction consists of two rectangles with random independent aspect ratios $r_1 \in [0.4, 3.6]$, $r_2 \in [0.5, 4.8]$ and rotation angle $\alpha \in [0, 2\pi]$ against each other. Finally landmark coordinates were randomly displaced by ± 1 pixelsize. The search procedure identifies the rotation point correctly. In Fig. 1 the aligned training data and the resulting first 3 modes from a single model are depicted. In the lower row the first modes of shape variation after automatic splitting

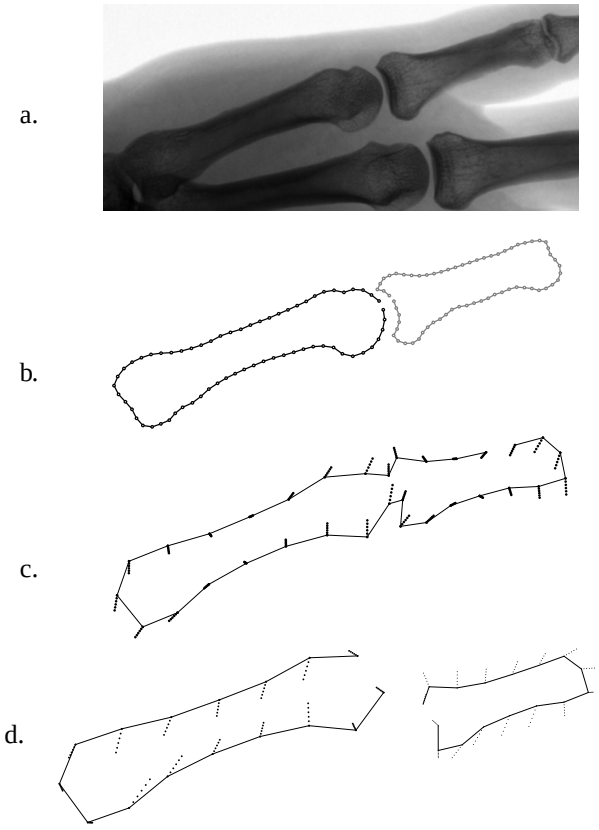

Figure 4: First modes of variation of the entire bone structure and of the two separate bones.

uncover the aspect ratio change. To represent 95% of the entire variation a single model needs 4 modes while with sub-shape models 2 modes are sufficient. The consequence will be explained later.

Bone data In a set of 40 radiographs *metacarpals 5* and *proximal phalanges 5* (Fig. 4a) were manually delineated by a radiologist. Landmark correspondences were established by the MDL based approach introduced in [7] and the search procedure was performed on 128 landmarks. Fig. 5 shows 16 hypotheses resulting from the growing procedure, Fig. 4b shows the winning hypothesis, splitting the structure at the rotation point between the two bones. The resulting first modes of variation for a single model and 2 sub-shape models are depicted in Fig. 4c and d, for clarity only 32 landmarks are shown. Note that since the radiographs were acquired during a standard examination procedure, the rotation variance is rather small, still it suffices to detect two distinct entities in the data. The reconstruction error using a model with 4 modes is 3.46 pixel for a single model versus 2.08 pixels with 2 sub-shape models. This allows for a more efficient search during application. The ratio between

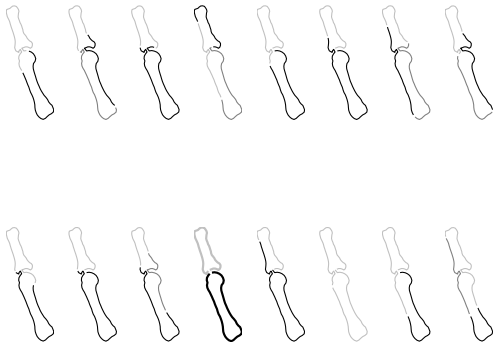

Figure 5: Set of hypotheses for the partitioning of bone shapes. Winning hypotheses is denoted with thick lines.

the criterion function of all possible partitions into two sub shape models and the criterion function for a single model as a function of the split positions is depicted in Fig. 3.

Hand contours For a set of 40 manually annotated hand contours with varying finger positions [15] our method was applied. First additional and more accurate correspondences on interpolated outlines were established again using MDL. A set of 16 hypotheses resulting from the growing on 128 landmarks is shown in Fig. 6, the final selection yields the winning hypothesis depicted in Fig. 7. The method nicely identifies the individual fingers as distinct sub-shape models.

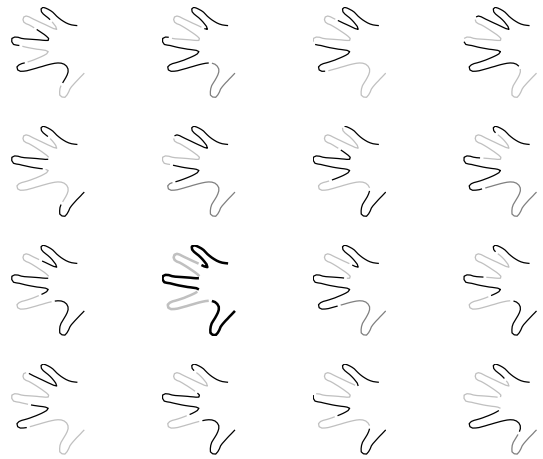

Figure 6: Set of hypotheses for the partitioning of hand shapes. Winning hypotheses is denoted with thick lines.

Compactness and reconstruction error For both a single model (dashed line) and automatically generated sub-shape models (solid line) in Fig. 8 the fraction of variance represented by a certain number of eigenvectors is depicted. In Fig. 9 the reconstruction error for a given number of used

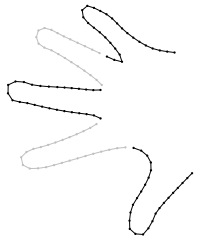

Figure 7: Winning Hypothesis of the hand data set.

eigenvectors is shown. For all experiments sub-shape models yield lower reconstruction error than single models, and the variation in the training data is more compactly represented in the first few eigenvectors. In Table 1 the number of modes necessary to represent 99% of the data variation are given. With optimal sub-shape models a smaller number is usually sufficient. These characteristics allow for the use of less modes during model application. Thus the search becomes computationally less expensive, and the generalization behavior improves with decreasing possibility of generating illegal shapes; i.e., the model fits better to the given data.

	Hands	Bones	Rectangles
sub-shape models:	4	5	4
single model:	8	8	6

Table 1: Number of modes necessary to represent 99% of the data variation.

4 Conclusion

In this paper an approach to automatically find optimal sub-shape models is proposed. Based on an MDL criterion data is divided into sub-shapes that allow for more compact modeling. The resulting model ensemble represents the variation present in the data in a more efficient way and thus lowers the computational complexity during model search.

Distinct entities are recognized without user interaction. They correspond fairly well to entities assigned by humans and pay respect to the capabilities of a certain model where manual partitioning can lead to sub-optimal results. In particular the method was evaluated with PCA based statistical shape models on three data sets where it resulted in more compact models. Ultimately the work aims at more and more autonomous model building processes, among others relevant in medical imaging. Ongoing work deals with application of the method to images for appearance based object recognition.

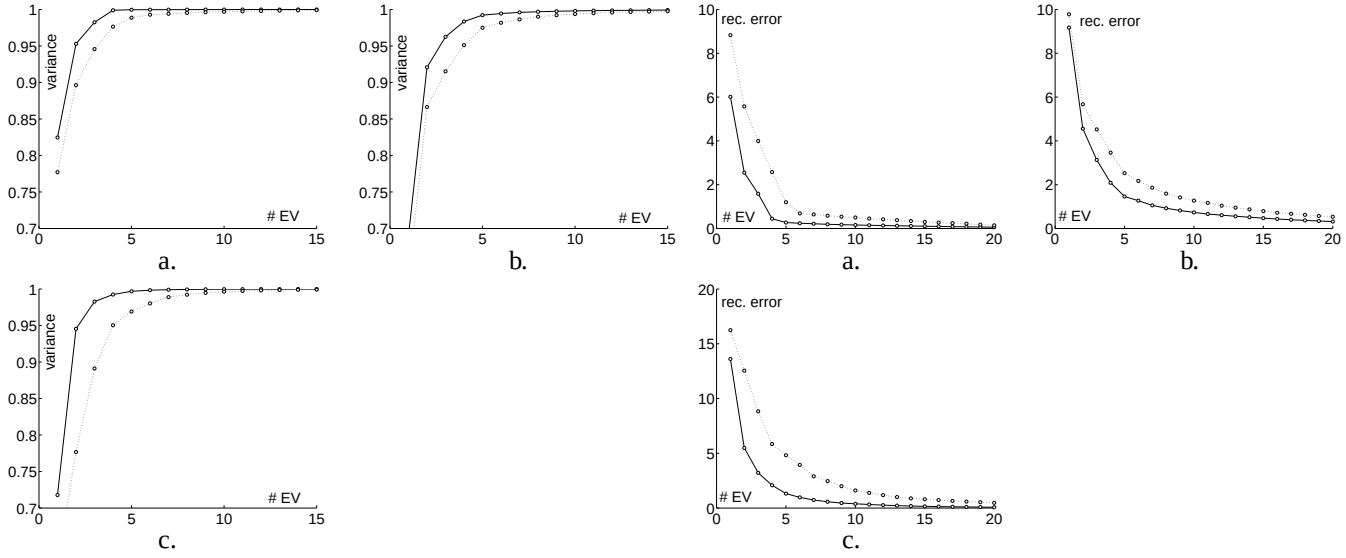

Figure 8: Captured variance for a. rectangle data b. bone data and c. hand contour data, the x-axis corresponds to the number of eigenvectors, single model dashed-line sub-shape model solid line.

Figure 9: Reconstruction error for a. rectangle data b. bone data and c. hand contour data, the x-axis corresponds to the number of eigenvectors, single model dashed-line sub-shape model solid line.

References

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proc. British Machine Vision Conference*, pages 266–275, Berlin, 1992. Springer.

[2] T. F. Cootes, A. Hill, Christopher J. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–366, 1994.

[3] Anonymous

[4] P.D.Sozou, T.F. Cootes, C.J. Taylor, and E.C.Di-Mauro. A non-linear generalisation of pdms using polynomial regression. In E. Hancock, editor, *Proceedings of the British Machine Vision Conference 1994, BMVC'94*, pages 397–406. University of York, BMVA, 1994.

[5] P.D.Sozou, T.F. Cootes, C.J. Taylor, and E.C.Di-Mauro. Non-linear point distribution modelling using a multi-layer perceptron. In D. Pycock, editor, *Proceedings of the British Machine Vision Conference 1995, BMVC'95*, pages 107–116. University of Birmingham, BMVA, 1995.

[6] T.F. Cootes and C.J. Taylor. A mixture model for representing shape variation. *Image and Vis. Comp.*, 17(8):567–574, 1999.

[7] Rhodri H. Davies, Carole j. Twining, Tim F. Cootes, John C. Waterton, and Chris J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21(5):525–537, May 2002.

[8] Rhodri H. Davies, Carole j. Twining, P. Daniel Allen, Tim F. Cootes, and Chris J. Taylor. Buidling optimal 2d statistical shape models. *Image and Vision ComputiNon-Linear Point Distribution Modelling using a Multi-Layer Perceptronng*, 21:1171–1182, 2003.

[9] C.J. Twining, S. Marsland, and C.J. Taylor. A unified information theoretic approach to the correspondence problem in image registration. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 3, pages 704–709, 2004.

[10] Aleš Leonardis and Horst Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000.

[11] Ales Leonardis, Horst Bischof, and Jasna Maver. Multiple eigenspaces. *Pattern Recogntion*, 35:2613–2627, 2002.

[12] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[13] J. Rissanen. Universal coding, information, prediction, and estimation. *IEEE Transactions on Information Theory*, 30:629–636, July 1984.

[14] C.E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 1948.

[15] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, March 2002.