A Clique of Active Appearance Models by Minimum Description Length *

Georg Langs[13], Philipp Peloschek[2], René Donner[13], Horst Bischof[1]

[1] Institute for Computer Graphics and Vision
Graz University of Technology, Austria
{langs, donner}@prip.tuwien.ac.at,
bischof@icg.tu-graz.ac.at

[2] Department of Diagnostic Radiology
Medical University of Vienna, Austria
philipp.peloschek@meduniwien.ac.at

[3] Pattern Recognition and Image Processing Group
Vienna University of Technology, Austria

Abstract

Autonomous model building is a crucial trend in model based methods like AAMs. This paper introduces an approach that deals with non-linearities by detecting distinct sub-parts in the data. Sub-models each representing an individual sub-part are derived from a minimum description length criterion. Thereby the resulting clique of models is more compact and obtains a better generalization behavior than a single model. The proposed AAM clique generation deals with non-linearities in the data in a generic information theoretic manner reducing the necessity of user interaction during training.

1 Introduction

Active Appearance Models (AAMs) [4] utilize principal component analysis for the generation of a linear model of shape and texture variation enabling an AAM search to detect objects even under difficult image conditions. They 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 [8].

Since non-linearities violate the linearity assumption of PCA utilized in AAM building, they degrade the compactness and thus the efficiency of the resulting models. Among other reasons non-linearities can occur due to movement of distinct anatomical structures. Various approaches to deal with specific non-linearities have been proposed. In [12, 13] they were dealt with by polynomial regression or multi-layer perceptrons. However the

*This research has been supported by the Austrian Science Fund (FWF) under the grant P17083-N04 (AAMIR). Part of this work has been carried out within the K-plus Competence center ADVANCED COMPUTER VISION funded under the K plus program.

BMVC 2005 doi:10.5244/C.19.87

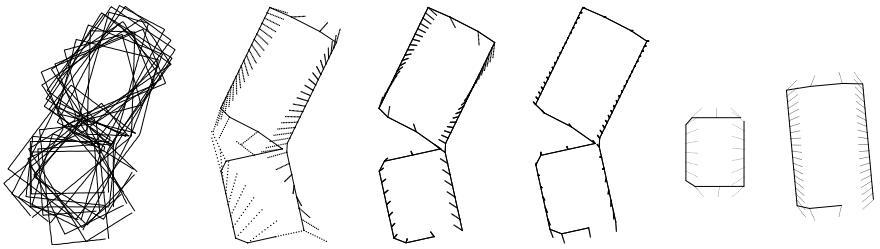


Figure 1: a.-d.: Aligned set of rotating rectangles, each with changing aspect ratio, and first 3 modes of variation. e.-f.: first modes for the separate rectangles (the small lines indicate the modes of variations for each landmark point).

order of polynomials or the architecture of the network had to be chosen application specific. In [3] mixture models were used resulting in more reliable models but becoming un-feasible for large training set sizes. In [8] snakes were used to deal with pathological local non-linearities during the search procedure. In Fig. 1 a simple example of non-linear shape variation is depicted. Two rectangles rotate against each other while independently changing aspect ratio. In Fig. 1a-d the aligned shape set and the first 3 modes of variation resulting from the entire shape are shown. 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 Fig. 1e-f the two rectangles are modelled separately and the change of aspect ratio is plausibly represented in the first modes. A correct determination of distinct entities in the data seems to be a worthwhile alternative to modelling the non-linear variations and is a crucial step to build compact and efficient models. In this paper the *minimum description length (MDL)* principles capability to perform an automatic identification of distinct entities will be explored.

The MDL principle has been used successfully as a model selection criterion in different applications. It allows for the comparison of likelihoods of different models that describe given data. In [9] an MDL based technique was used for image segmentation. In [5] MDL was used to establish landmark correspondences on a set of shapes defined by continuous contours before active shape model training was performed in order to obtain a model not compromised by artefacts of mere landmark placement. In [17] group-wise non-rigid registration was performed with help of MDL. A method proposed in [11] uses MDL to select hypotheses for robust appearance based object recognition. In [10] 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 contribution of this paper is an algorithm that splits AAMs in order to model given training data with a set of sub-models (we call this set *model clique*) instead of a single model. A general criterion function based on the MDL principle is proposed making the approach applicable to other models as well. The AAM clique generation procedure utilizes this criterion for the splitting of AAMs with respect to the eigenspace model representing the landmarks. Finally building a multiple AAM clique taking texture into account and allowing for efficient search is explained. The work is an extension

of optimal sub-shape model generation [7] where connected shape data is split without dealing with texture or disconnected sub-models.

The paper is structured as follows: In Sec. 2 a criterion function for multiple model selection is introduced, which in Sec. 3.1 is used to determine an optimal sub division of given data. In Sec. 3.2 the composition of the corresponding AAM clique is described. After experimental results are presented in Sec. 4, in Sec. 5 a final discussion is given.

2 A criterion for multiple model selection

After proposing an MDL formulation for multiple models a criterion function that allows for the splitting of eigenspace shape models will be formulated.

Minimum description length In order to find an optimal model clique describing the data set the minimum description length principle is used [14, 15]. It states that maximizing the likelihood of a model $\mathscr{M}$ 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.

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

To model the AAM shape data PCA is applied to the coordinate vectors defining the positions of the landmarks. The MDL criterion will be used to judge the encoding of the shape data with cliques of sub AAMs, and ultimately aims at obtaining an optimal sub-division of the data based on the spatial variation.

Multiple models A set of models $\{\mathscr{M}_1,\ldots,\mathscr{M}_n\}$ 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 clique* $\mathsf{M} = \langle \mathscr{M}_1,\ldots,\mathscr{M}_n;\mathscr{S}\rangle$, where $\mathscr{S}$ holds the information of the data parts corresponding to the individual sub-models. We minimize

$$C(\mathsf{M}) = L(\mathscr{S}) + \sum_{\mathscr{M}_i\in\mathsf{M}} L(\mathscr{M}_i) + L(D_i|\mathscr{M}_i) + \mathscr{R}, \tag{2}$$

where $L(\mathscr{S})$ is the additional cost for transmitting the splitting information, and $\mathscr{R}$ is a penalty for the residual error. This corresponds to the maximization of the likelihood of the model clique.

Description length of statistical shape models AAM represent shapes by a finite set of n landmarks. Each of n_T shapes in the training set can then 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 creates a new coordinate system that represents each of the vectors

$$\mathbf{x}_i = \overline{\mathbf{x}} + \sum_{j=1}^{n_p} a_j \mathbf{e}_j, \tag{3}$$

in an optimal way. The modes $\mathbf{e}_j$ are the eigenvectors of the covariance matrix sorted according to decreasing eigenvalue λ_j. $\overline{\mathbf{x}}$ 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 $\mathbf{e}_j$.

If we model shape data by a multivariate Gaussian in the directions of the decorrelated eigenvectors $\mathbf{e}_j$ as described above we can apply Shannons coding theorem [16] 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 $\hat{a}_j^i = k\Delta_{Im}, k \in \mathbf{Z}$ with $-R_j/2 \leq \hat{a}_j \leq R_j/2$ are modelled 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 $L(\mathscr{M}_{\mathbf{e_j}})$ are the quantized eigenvector, $\hat{\sigma}_j$ and the quantization parameter δ_j for the direction $\mathbf{e}_j$. $L(D|\mathscr{M}_{\mathbf{e_j}})$ 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 penalty for the residual error

$$\sum_{j=1}^{n_p} \Big(L(\mathscr{M}_{\mathbf{e_j}}) + L(D|\mathscr{M}_{\mathbf{e_j}}) \Big) + \mathscr{R}. \tag{4}$$

The Criterion function for muliple eigenspace models A set of n_T example shapes each defined by n corresponding landmarks is given. The general MDL formulation in Eq. 2 can now be applied to a set of multivariate Gaussians each representing a sub-shape i.e. a sub set of these landmarks. From that we will derive the criterion function allowing for automatic splitting of AAMs. Although the primary thread refers to shape models the application to any vector data is straightforward. When Eq. 2 is applied to a clique M of eigenspace models $\mathscr{M}_m$ the criterion becomes

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

where $n_p^m = \max\{j : \sigma_j > \Delta_{Im}\}$ is the dimension of the utilized eigenspace for $\mathscr{M}_m$. $L(\mathscr{S})$ is the additional cost to transmit the split information. This term acts as a penalty for additional splits and prohibits possible trivial solutions. In our case $L(\mathscr{S})$ is the cost of assigning each of n landmarks a sub-model. Assuming $l \geq 1$ sub-models and equal probability for all possible split positions, the cost is

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

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

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

where $\sigma_{max} = R/2$ and $\sigma_{min} = 2\Delta_{Im}$. R is the strict upper bound of the coefficients w.r.t. the distribution. A detailed derivation of the description length of 1D Gaussians is given in [5]. $\mathscr{R}$ is the penalty for the residual error that remains after fitting the training set with the model

$$\mathscr{R} = n_T \sum_{j=n_p+1}^{2n} \lambda_j. \tag{8}$$

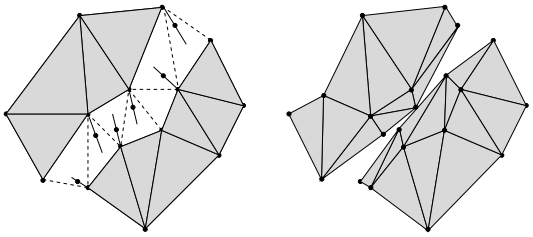


Figure 2: Generation of new texture patches (gray); left: initial triangulation, the dashed lines are edges that now connect different sub-models; right: triangulation after adding additional landmarks

3 Splitting AAMs

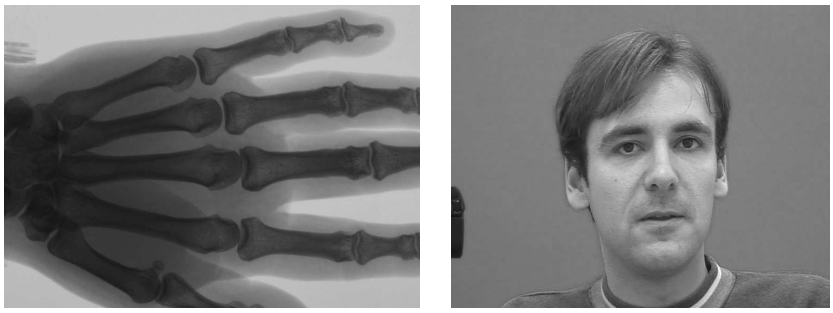
Utilizing Eq. 5 an optimal sub division of an AAM can be accomplished by optimization. An AAM clique representing the training data in a more efficient manner can be generated by splitting the AAM based on the shape information. Taking texture into account would make a deformation of the entire training set texture with respect to the landmark sub-division necessary for each step, making it unfeasible for optimization. However the landmark information is sufficient to give a sub-division that enhances texture representation as well (see Fig. 5c).

3.1 Finding an optimal sub-model clique

Given a set of shapes, the criterion in Eq. 5 is used to find an optimal set of sub-models describing the training set in a more efficient manner than a single model would do. Although constraints like connectivity conditions or a priori neighbourhood relations could be used to influence the sub-model generation they are not necessary. The results in this paper are derived only from the landmark coordinate information without constraints enabling the algorithm to generate non-connected sub-models which is sensible in cases where symmetry is present e.g. faces. For the clarity of presentation additional constraints were set aside during the experiments. However using these contraints a considerable speed up might be obtained. The algorithm to generate the sub-model clique performs as follows:

1. Initialization Given n_T sets of n corresponding landmarks $\{\mathbf{x}_i\}_{i=1,\ldots,n}$ and a number L of sub-models, each landmark is assigned a random sub-model label $j \rightarrow [1,L]$ resulting in a membership function $\mathbf{m} \in \{1,\ldots,L\}^n$.

2. Optimization Each data part $\{\mathbf{x}_i^l\}_{i=1,\ldots,n_l}$ corresponding to a sub-model l i.e. where $\mathbf{x}_i^l = \langle x_j | \mathbf{m}(j) = l \rangle$ is aligned, the eigenspace of shape variation is calculated and the description length of the entire clique is determined according to Eq. 5. A simulated annealing process [6] optimizes $\mathbf{m}$ with respect to $C(\mathsf{M})$. That is in each step a landmark can change its sub-model i.e. j_r and $l_r \in \{1,\ldots,L\} \setminus \mathbf{m}(j_r)$ are chosen randomly, $\mathbf{m}'(j_r) = l_r$ and the corresponding new criterion function value $C(\mathsf{M}')$ is calculated. Given



a. b.

Figure 3: a. Medical radiograph data and b. face video data.

a temperature value T the membership function is updated $\mathbf{m} \rightarrow \mathbf{m}'$ with probability $p = exp(\delta(C(\mathsf{M}') - C(\mathsf{M}))/T)$. With a reasonable decrease schedule of T during the process it converges at the global minimum of $C(\mathsf{M})$ with high probability, resulting in the final model clique M_{final}.

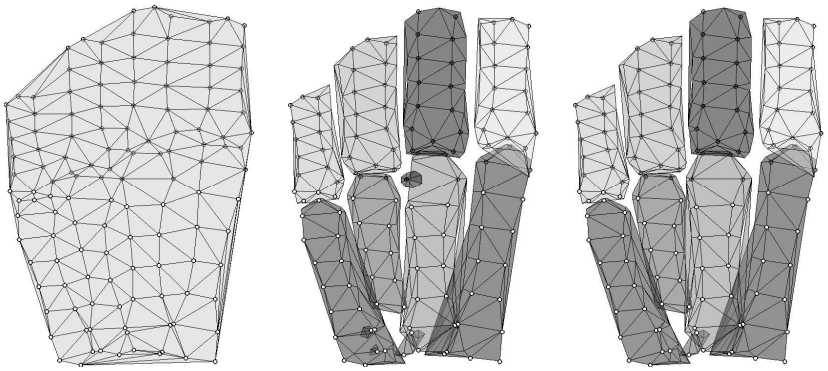
3. Cleaning Although from a purely information theoretic standpoint M_{final} has to be considered as optimal, a final step using a neighbourhood relation $N(i,j)$ established by a simple Delaunay triangulation of the mean shape can improve the result slightly. The use of neighbourhood is a reasonable due to the spatial nature of the data. Hence during a final cleaning step isolated landmarks i.e. landmarks $l_{iso} \in \mathscr{M}_i$ that are totally enclosed by different sub-models $l_n \in \mathscr{M}_{j_n}, j_n \neq i$ change their membership to the sub-model $\mathscr{M}_k$ with the highest number of neighbours to l_{iso}. This procedure results in $\mathsf{M}'_{\mathsf{final}}$.

3.2 Composing the final AAM clique

The model clique $\mathsf{M}_{\mathsf{final}}$ resulting from the opimization gives a partitioning of the set of landmarks constituting the single AAM. Each of the sub-models exhibits its one parameter set. A straightforward Delaunay triangulation of each landmark sub set could result in sub optimal texture representations due to gaps between sub-models and large overlapping texture patches. In order to represent the texture properly the landmark triangulation is performed as follows:

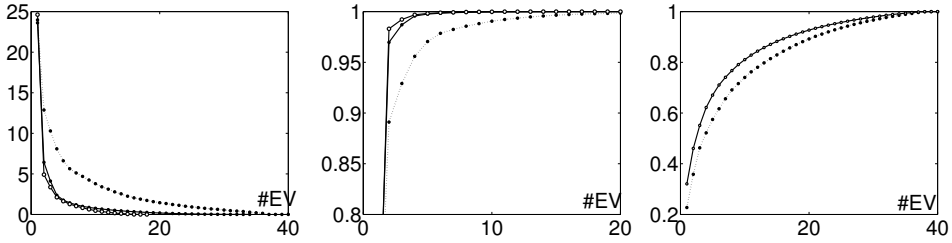
1. Adding additional landmarks Given the sub-model landmarks $\mathbf{x}_i^l$ parts that were not connected in the initial Delaunay triangulation $\mathbf{tri}_{single}$ of the single model stay non-connected. For each of these parts a set of additional landmarks is generated in order to capture texture information in between sub-AAMs. In Fig. 2 two parts of different sub-models are depicted. For each landmark $s_1 \in \mathscr{M}_l$ that is corner of a triangle $tri_i \in \mathbf{tri}_{single}$ with 2 landmarks s_2, s_3 part of a different sub-model an additional landmark $s_{add} \in \tilde{\mathscr{M}}_l$ is placed on the ray defined by the landmark and the centroid of the triangle. The distance is defined by a factor d_{add} determing how much of the texture between the sub-AAMs is to be represented. Since the additional landmarks depend only on one sub-model a seamless coverage of the texture between the sub-models is no longer ensured. If inbetween texture is important an overlap can be effected by increasing d_{add}.

2. Sub-model triangulation Each part of the landmarks of the resulting sub models $\tilde{\mathbf{x}}_i^l$ is Delaunay triangulated within its convex hull. However if necessary a more restrictive



a. b. c.

Figure 4: a. Initial data, b. generated AAM clique and c. cleaned AAM clique for hand radiograph data.



a. b. c.

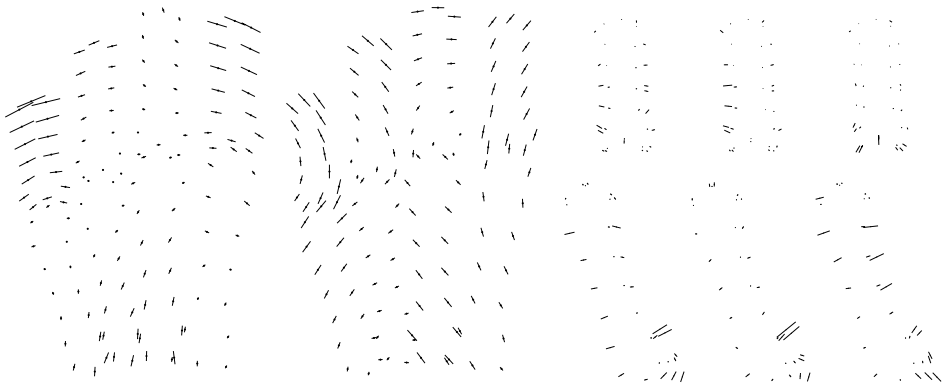
Figure 5: a. Reconstruction error, b. captured variance for landmarks and c. for texture of hand bones. dashed line: single model, solid line: model clique, circles: cleaned model clique.

hull can be chosen in order to avoid a high degree of overlapping texture patches.

4 Experiments

Setup Evaluation results will be given for two data sets: 1. For 40 hand radiographs metacarpals and proximal phalanges $2-5$ were segmented by a radiologist and correspondences for 128 landmarks were established by an MDL based method [5]. 2. Face data of a face talking made available by [1]. 68 landmarks were tracked by an AAM over a sequence of 5000 frames of which 80 frames were randomly picked for model building and splitting. The resulting AAM cliques were evalutated with respect to the reconstruction error and the variation captured in the model.

Hand data In Fig. 4 the splitting results for the medical data is presented. Fig. 4a. shows the texture triangulation of a single AAM, Fig. 4b. shows texture triangulations for an AAM clique composed of 8 sub-models resulting from simulated annealing and Fig. 4c. shows the AAM clique after cleaning. The decomposition of anatomical structures is nearly correct. Although the radiographs were acquired during a standard protocol



a. b. c.

Figure 6: Bone data: a. 1^{st} and b. 2^{nd} mode of shape variation for single model, c. 1^{st},2^{nd} and 3^{rd} mode of shape variation for sub models.

designed to avoid varying hand posture i.e. individual bone rotations even the bones in the carpus are detected as distinct entitites. In the proximal region overlaps of the bones occur and are mirrored in the resulting sub-models. Fig. 5 gives a quantitative evaluation of the resulting AAM clique. With 3 modes the AAM clique represents 98.7% of the variation in the data as opposed to 89.1% with a single model. The cleaning step improves this compactness only marginally to 99.2%. The reconstruction error with 5 modes lies at 1.6 pixel for the model clique, and at 6.6 pixel for the single model. Even though only landmark information was used during the clique generation the compactness of the texture model improves as well (Fig. 5c). Figs. 6 and 7 depict the first modes of variation for the single model and two sub-models for shape and texture, respectively. It can be observed that the variation of bone shape is represented much more explicitly by the sub-models.

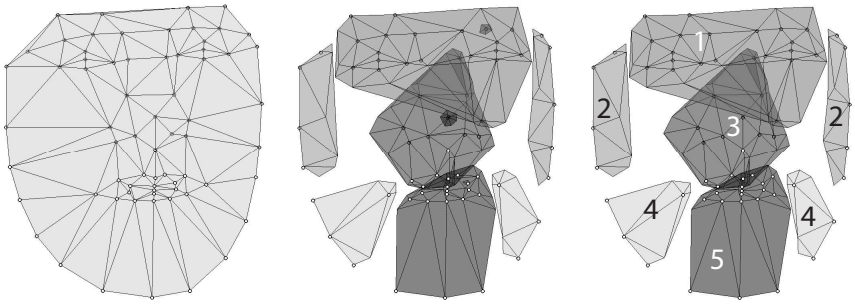
Face Data Fig. 8 shows the splitting results for the data of 80 frames taken from a sequence of a talking face. Sub-models are indicated by numbers. The AAM clique is constituted of 5 sub-models situated symmetrically w.r.t. the vertical face axis. 2 of the sub-models consist of non-connected parts. The quantitative evaluation indicates only minor improvement of the model compactness, for 3 modes represented variation is increased from 98.3% to 99.3%. The reason for the small improvement could lie in the data acquisition by an AAM based tracking method. Nevertheless the automatic partition of the face landmarks corresponds to intuitively sensible parts of a face (eyes, nose, mouth) and mirrors the inherent symmetry soundly. A connectivity constraint would have hindered this result. An interesting obeservation is that although no connectivity or neighbourhood constraints are applied large connected components of spatially neighbouring landmarks emerge fairly early in the optimization process.

5 Conclusion

This paper proposes an MDL based method to automatically split active appearance models into cliques of sub-models. Thereby the compactness of the resulting representation can be increased resulting in improved generalisation behavior and a more efficient AAM

a. b. c.

Figure 7: Bone data: a. 1^{st} and b. 2^{nd} mode of texture variation for single model, c. 1^{st},2^{nd} and 3^{rd} mode of texture variation for sub models.



a. b. c.

Figure 8: Face data: a. Initial data, b. generated AAM clique and c. cleaned AAM clique.

search. Results indicate that the method is able to distinguish anatomical structures or even facial features by using their modelling behavior. It deals with non-linearities in the data in a generic manner allowing for less user interaction during training.

The optimization algorithm utilized is applicable to other data as well. The interdependency between the sub-models are neglected so far, current work concentrates on a feasible meta-model and an optimization procedure including assignment of the number of necessary sub-models. Future work will concentrate on efficient search methods for cliques of sub-models and application to other data not necessarily of spatial nature. Ultimatly this work aims at more autonomous model building techniques being able to cope with complex data considering the occurence of distinct entities.

References

[1] FGnet - IST-2000-26434, Face and Gesture Recognition Working Group.

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

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

[4] T.F. Cootes, J. Edwards, and C.J. Taylor. Active appearance models. *IEEE Trans. PAMI*, 23(6):681–685, 2001.

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

[6] S. Kirkpatrick, C.D. Gerlatt Jr., and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[7] G. Langs, P. Peloschek, and H. Bischof. Optimal sub-shape models by minimum description length. In *Proceedings of CVPR 2005*, vol. 2, pages 310–315, 2005.

[8] G. Langs, P. Peloschek, and H. Bischof. ASM driven snakes in rheumatoid arthritis assessment. In *Proceedings of SCIA 2003*, LNCS 2749, pages 454–461.

[9] Y.G. Leclerc. Image segmentation via minimal-length encoding. In *Proceedings of the 6th Multidimensional Signal Processing Workshop, IEEE Acoustics, Speech and Signal Processing Csociety*, page 78, 1989.

[10] A. Leonardis, H. Bischof, and J. Maver. Multiple eigenspaces. *Pattern Recogntion*, 35:2613–2627, 2002.

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

[12] 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 BMVC'94*, pages 397–406. University of York, BMVA, 1994.

[13] 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 BMVC'95*, pages 107–116. University of Birmingham, BMVA, 1995.

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

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

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

[17] C.J. Twining, S. Marsland, and C.J. Taylor. A unified information theoretic approach to the correspondence problem in image registration. In *Proceedings of ICPR 04*, volume 3, pages 704–709, 2004.