

Technical University of Denmark



Elastic Appearance Models

Hansen, Mads Fogtmann; Fagertun, Jens; Larsen, Rasmus

Published in:

Proceedings of the British Machine Vision Conference

Link to article, DOI:

[10.5244/C.25.91](https://doi.org/10.5244/C.25.91)

Publication date:

2011

Document Version

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):

Hansen, M. F., Fagertun, J., & Larsen, R. (2011). Elastic Appearance Models. In Proceedings of the British Machine Vision Conference BMVA Press. DOI: 10.5244/C.25.91

DTU Library

Technical Information Center of Denmark

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Elastic Appearance Models

Mads Fogtmann Hansen
<http://www.imm.dtu.dk/~mfh>

Jens Fagertun
<http://www.imm.dtu.dk/~jenf>

Rasmus Larsen
<http://www.imm.dtu.dk/~rl>

DTU Informatic
Technical University of Denmark
Kgs. Lyngby, Denmark

Abstract

This paper presents a fusion of the active appearance model (AAM) and the Riemannian elasticity framework which yields a non-linear shape model and a linear texture model – the active elastic appearance model (EAM). The non-linear elasticity shape model is more flexible than the usual linear subspace model, and it is therefore able to capture more complex shape variations. Local rotation and translation invariance are the primary explanation for the additional flexibility. In addition, we introduce global scale invariance into the Riemannian elasticity framework which together with the local translation and rotation invariances eliminate the need for separate pose estimation. The new approach was tested against AAM in three experiments; face labeling, face labeling with poor initialization and corpus callosum segmentation. In all the examples the EAM performed significantly better than AAM. Our Matlab implementation can be downloaded through svn from <https://svn.imm.dtu.dk/AAMLab/svn/AAMLab/trunk/>

1 Introduction

Generative model-based image interpretation for image segmentation, registration and labeling plays a central role in many image analysis applications. One of the most successful model-based approaches is the *Active Appearance Model* by Cootes *et al.* [6] which applies learned shape and texture (intensity) priors from a training set. The method has been applied to medical image segmentation [9, 9, 17], eye tracking [11, 14], object tracking [22] and facial recognition [10, 9, 10] with success.

AAM creates a shape model from set of training shapes by modeling them in a linear subspace. Linear generative shape models will in general not be able to capture complex shape variations as biological shapes rarely can be embedded in a linear subspace. This may result in folding shapes and poor registration/segmentation. Instead, we will assign an individual displacement to each point in the mean shape and construct a statistical local deformation prior which measures how well a given local deformation matches the local deformations in the training set. We use the Riemannian elasticity framework [18] to model this prior. Riemannian elasticity priors can capture complex deformations as they are locally rotation and translation invariant.

2 Previous work

Many improvements to AAM have been proposed to strengthen the approach with respect to efficiency, discrimination and robustness, see Gao *et al.* [8] for a complete survey. Many improvements deal with reducing the computational cost by tuning the optimization, e.g. the inverse compositional algorithm by Baker *et al.* [2], or by compressing the texture model, e.g. the wavelet and wedgelet model by Larsen *et al.* [13]. Other research groups have increased the robustness of AAM by combining it with the active shape model [23] or by making it light invariant [12]. Suggestions to improve the accuracy of the AAM include the maximum a posteriori formulation by Cootes *et al.* [6] and inclusion of local image structure by Scott *et al.* [20].

To our knowledge Christoudias and Darrell [4] are the only researchers to propose non-linear shape modeling extensions to AAM. There are complementary methods to AAM who use non-linear shape modeling, e.g. the deformable M-rep [19].

3 Linear appearance modeling

In AAM the goal is to construct a compact parametrization of the variability of an object, described by its shape and texture, from a set of (semi-)manually labeled images. Generalized Procrustes analysis filters out rotation, translation and scaling effects from the landmarks yielding a mean shape and a set of aligned shapes. A linear shape model follows from a truncated matrix decomposition, e.g. Principal Components Analysis, on the aligned shapes, i.e.

$$\mathbf{v} = \bar{\mathbf{v}} + \mathbf{U}_v \mathbf{p}_v, \quad (1)$$

where $\bar{\mathbf{v}}$ is the mean shape and \mathbf{U}_v contains the modes of shape variation. Warping the textures to a common reference space, defined by the mean shape, creates shape-free textures, and another truncated matrix decomposition generates a linear texture model, i.e.

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{U}_g \mathbf{p}_g. \quad (2)$$

A final truncated matrix decomposition on the combined set of shape and texture parameters yields a combined appearance model

$$\begin{bmatrix} \mathbf{v} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{v}} \\ \bar{\mathbf{g}} \end{bmatrix} + \begin{bmatrix} w\mathbf{U}_v & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_g \end{bmatrix} \mathbf{U} \mathbf{p}, \quad (3)$$

where w weights the relative importance between shape and texture.

4 Non-linear shape modeling using elasticity

Linear generative shape models, Eqn. 1 and Eqn. 3, will in general not be able to capture complex shape variations. This section introduces non-linear strain modeling and describes how a Riemannian elasticity prior can be constructed.

4.1 Representing local deformation

Consider the particle P and the neighboring particle Q in an elastic body positioned at \mathbf{x} and $\mathbf{x} + d\mathbf{x}$, respectively (see Figure 1(a)). The elastic body undergoes a diffeomorphic

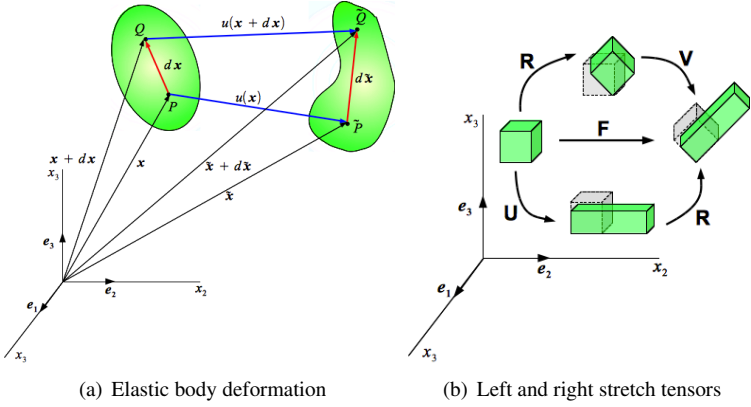


Figure 1: Representing local deformation.

deformation, $\phi(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$, such that the particle P and Q move to positions \tilde{P} and \tilde{Q} given by $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}} + d\tilde{\mathbf{x}}$. Obviously, there exists a linear mapping between $d\mathbf{x}$ and $d\tilde{\mathbf{x}}$, i.e. $d\tilde{\mathbf{x}} = \mathbf{M}d\mathbf{x}$, which represents the local transformation. From Figure 1(a) it can be derived that

$$d\tilde{\mathbf{x}} = d\mathbf{x} + \mathbf{u}(\mathbf{x} + d\mathbf{x}) - \mathbf{u}(\mathbf{x}). \quad (4)$$

By performing a first-order Taylor expansion of $\mathbf{u}(\mathbf{x} + d\mathbf{x})$ around P we get

$$d\tilde{\mathbf{x}} \approx d\mathbf{x} + \mathbf{u}(\mathbf{x}) + \nabla\mathbf{u}(\mathbf{x})d\mathbf{x} - \mathbf{u}(\mathbf{x}) = (\mathbf{I} + \nabla\mathbf{u}(\mathbf{x}))d\mathbf{x} = \mathbf{F}(\mathbf{x})d\mathbf{x} \quad (5)$$

Thus, for a sufficiently small $d\mathbf{x}$ we get the relationship $\mathbf{M} \approx \nabla\phi(\mathbf{x}) = \mathbf{F}(\mathbf{x})$.

As the deformation gradient tensor \mathbf{F} is based on the displacement gradient it is translation invariant but not rotation invariant. However, \mathbf{F} can be decomposed into an orthonormal tensor (representing the rotation) and a positive definite symmetric tensor (representing the stretch), i.e.

$$\mathbf{F} = \mathbf{R}\mathbf{U} = \mathbf{V}\mathbf{R}. \quad (6)$$

\mathbf{U} and \mathbf{V} are denoted the *right stretch tensor* and the *left stretch tensor* respectively. \mathbf{U} and \mathbf{V} have identical eigenvalues or *principal stretches* but different eigenvectors. The decomposition is illustrated in Figure 1(b).

From the right stretch tensor we define a sequence of rotation and translation invariant tensors which are known as the *Lagrangian strain tensors*

$$\mathbf{E}_m = \frac{1}{2m}(\mathbf{U}^{(2m)} - \mathbf{I}). \quad (7)$$

These tensors essentially model the deviation of the right stretch tensor from the identity. The special case $\mathbf{E}_0 = \log(\mathbf{U})$ is called the *Hencky strain tensor* which will be the basis for our non-linear shape model.

4.2 Statistical deformation modeling

Given the Lagrangian strain tensors it is possible to formulate a rotation and translation invariant density function which quantifies the energy stored in an infinitesimal neighborhood due to the deformation. The most intuitive way to build a strain energy density function is

just to use the magnitude, defined by a suitable matrix norm (e.g. the Frobenius norm), of a suitable strain tensor. To protect against folding we want an energy density function, which approaches infinity when the approaches towards the black-hole deformation. This implies that the at least one of the eigenvalues of the strain tensor should go to infinity when the determinant of the Jacobian of the deformation gradient approaches zero. The Hencky strain tensor, \mathbf{E}_0 , fulfils this requirement. Pennec *et al.* [18] proposed the *statistical Riemannian elasticity energy*

$$W(\mathbf{E}_0) = \begin{cases} \frac{1}{4} \|\text{vect}(\mathbf{E}_0 - \bar{\mathbf{E}}_0)\|_{\Sigma^{-1}}^2 & \text{if } \det(\mathbf{F}) > 0 \\ \infty & \text{otherwise.} \end{cases} \quad (8)$$

where Σ and $\bar{\mathbf{E}}_0$ are calculated from a set of n previously observed deformations. Hence,

$$\bar{\mathbf{E}}_0 = \frac{1}{n} \sum_{i=1}^n \mathbf{E}_{0,i}, \text{ and } \Sigma = \frac{1}{n} \sum_{i=1}^n \text{vect}(\mathbf{E}_{0,i}) \text{vect}(\mathbf{E}_{0,i})^T. \quad (9)$$

5 Material and Methods

5.1 Warping and discretization

The deformation model, or the warp model as it is traditionally coined in AAM, enables us to separate the shape and texture information of a given image by *warping* the texture to the mean shape. There are multiple ways of defining a warp using the points of the mean shape and the deformed shape as control points. However, it is standard practice to triangulate the mean shape and use linear interpolation to propagate vertex displacements to the points inside the simplexes. EAM adapts this approach.

The affine transformation between a simplex $\{\mathbf{v}_1, \dots, \mathbf{v}_{N+1}\}$ and a simplex $\{\mathbf{v}_1 + \Delta\mathbf{v}_1, \dots, \mathbf{v}_{N+1} + \Delta\mathbf{v}_{N+1}\}$ can be defined by assuming that the point \mathbf{x} inside $\{\mathbf{v}_1, \dots, \mathbf{v}_{N+1}\}$ corresponds to the point $\phi(\mathbf{x})$ in $\{\mathbf{v}_1 + \Delta\mathbf{v}_1, \dots, \mathbf{v}_{N+1} + \Delta\mathbf{v}_{N+1}\}$ which has the same barycentric coordinates, see Figure 2. The barycentric coordinates, $\boldsymbol{\eta}$, of an arbitrary point, \mathbf{x} , inside the simplex $\{\mathbf{v}_1, \dots, \mathbf{v}_{N+1}\}$ are then given by

$$\begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_{N+1} \\ 1 & \cdots & 1 \end{bmatrix} \boldsymbol{\eta} = \begin{bmatrix} \mathbf{V} \\ \mathbf{1}^T \end{bmatrix} \boldsymbol{\eta}, \boldsymbol{\eta} = \begin{bmatrix} \mathbf{V} \\ \mathbf{1}^T \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = [\mathbf{D} \quad \boldsymbol{\eta}^0] \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}, \quad (10)$$

where $\boldsymbol{\eta}^0$ is the barycentric coordinates for origo. Thus, the transformation between two simplices is given by

$$\phi(\mathbf{x}) = \mathbf{x} + \Delta\mathbf{V}\boldsymbol{\eta} = \mathbf{x} + \Delta\mathbf{V} [\mathbf{D} \quad \boldsymbol{\eta}^0] \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}, \nabla\phi(\mathbf{x}) = \mathbf{I} + \Delta\mathbf{V}\mathbf{D}. \quad (11)$$

Note, as the transformation is linear the spatial derivatives of the transformation are independent of the spatial position within the fixed simplex. This means that the density energy in Eqn. 8 is identical in all points within a simplex, and thus we only have to estimate Σ and $\bar{\mathbf{E}}_0$ once for every simplex instead of once for every pixel position in the shape-free reference space. The integrated elasticity energy for the entire deformation is

$$r(\Delta\mathbf{v}) = \sum_i^n a_i W_i (\log((\mathbf{I} + \Delta\mathbf{V}_i \mathbf{D}_i)^T (\mathbf{I} + \Delta\mathbf{V}_i \mathbf{D}_i))), \quad a_i = \frac{1}{2} \left| \det \left(\begin{bmatrix} \mathbf{V}_i \\ \mathbf{1}^T \end{bmatrix}^{-1} \right) \right|, \quad (12)$$

where n is the number of simplices.

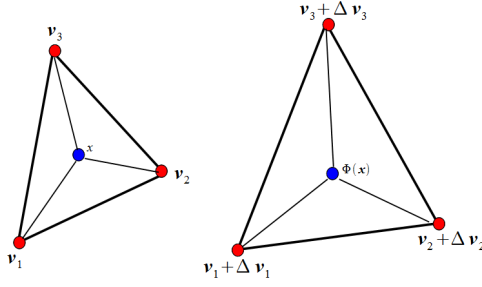


Figure 2: Illustration of the affine transformation $\Phi(\mathbf{x})$ from triangle (3-simplex) $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ to triangle $\{\mathbf{v}_1 + \Delta\mathbf{v}_1, \mathbf{v}_2 + \Delta\mathbf{v}_2, \mathbf{v}_3 + \Delta\mathbf{v}_3\}$.

5.2 Active appearance model

In AAM, the model is matched to a template image, \mathbf{T} , by (see [8])

$$\operatorname{argmin}_{\mathbf{p}} \|\mathbf{g}(\mathbf{p}) - \mathbf{T} \circ \phi(\mathbf{p})\|^2 + \alpha_1 \|\mathbf{p}\|_{C^{-1}}^2, \quad (13)$$

where C is the covariance matrix of \mathbf{p} and α_1 controls the strength of the parameter prior. α_1 is often set to zero and replaced by lower and upper bounds on \mathbf{p} . As it, in practice, is impossible to initialize the matching perfectly with respect to pose, pose estimation is added to the optimization problem, i.e.

$$\operatorname{argmin}_{\mathbf{p}, \mathbf{t}} \|\mathbf{g}(\mathbf{p}) - \mathbf{T} \circ \boldsymbol{\omega}(\mathbf{t}) \circ \phi(\mathbf{p})\|^2 + \alpha_1 \|\mathbf{p}\|_{C^{-1}}^2, \quad (14)$$

where $\boldsymbol{\omega}$ is the similarity transformation.

5.3 Elastic appearance model

Similar to AAM, the EAM matches the appearance model to the template image using the sum of squared residuals with the addition of the statistical elasticity energy, i.e.

$$\operatorname{argmin}_{\mathbf{p}, \Delta\mathbf{v}} \|\mathbf{g}(\mathbf{p}) - \mathbf{T} \circ \phi(\Delta\mathbf{v})\|^2 + \alpha_1 \|\mathbf{p}\|_{C^{-1}}^2 + \alpha_2 r(\Delta\mathbf{v}). \quad (15)$$

The advantages of Eqn. 15 compared to Eqn. 13-14 are that it cannot produce folding warps, and that it is not necessary to optimize rotation and translation separately due to local invariance. The disadvantage is that it is no longer a least square problem and thus requires the application of more advanced optimization algorithms. Our experience is that a limited-memory BFGS optimizer works well with the problem. Eqn. 15 is not scale invariant which implies that Eqn. 15 cannot match the shape model to an image if the scale between them is too large or small. One way to handle this is to introduce a global scaling parameter into the optimization. To avoid introducing a new parameter we prefer to filter out global scaling effects from the shape displacements prior to calling the regularizer, i.e.

$$\operatorname{argmin}_{\mathbf{p}, \Delta\mathbf{v}} \|\mathbf{g}(\mathbf{p}) - \mathbf{T} \circ \phi(\Delta\mathbf{v})\|^2 + \alpha_1 \|\mathbf{p}\|_{C^{-1}}^2 + \alpha_2 r\left(\frac{\Omega_T(\bar{\mathbf{v}})}{\Omega_T(\bar{\mathbf{v}} + \Delta\mathbf{v})}(\bar{\mathbf{v}} + \Delta\mathbf{v}) - \bar{\mathbf{v}}\right), \quad (16)$$

where $\Omega_T(\mathbf{v})$ computes the area of a shape defined by the triangulation T and the vertices \mathbf{v} .

Even though Eqn. 16 eliminates the need for separate pose estimation there are situations in practise where it is an advantage to restrict the displacement update to that of a similarity transformation – especially when the initial pose may be far from the true. On the other hand, if you are tracking an object frame-by-frame there is no need for such restrictions.

5.4 Implementation details

The AAM and EAM algorithms were implemented in Matlab using object oriented programming. Both the AAM og EAM uses a limited-memory BFGS optimizer for the model matching. The AAM implementation also contains the traditional *linear relationship* optimization strategy proposed by Cootes *et al.* [6] but it was not used for the experiments in this paper as it in general performed inferior to the BFGS optimization. The minFunc and minConf optimization packages by Schmidt[24] is used in our implementation. Our implementation can be downloaded from <https://svn.imm.dtu.dk/AAMLab/svn/AAMLab/trunk/> using a svn client with username `guest` and password `aamlab`. Documentation can be found at <http://www2.imm.dtu.dk/~mfh/AAMLAB/>.

5.5 Experiments

The performance of the EAM was tested against the AAM in three experiments.

1) *Face labeling*: The capability of the AAM and EAM to model and label faces. This experiment used the subset of neutral facial expression images in the AR database [15] where annotation markup data were available from FGnet – in total 119 images. The data set images have a resolution of 768×576 with the mean distance of 111.2 pixels between eyes of the recorded persons. The performance of the methods was evaluated using leave-one-out cross-validation with the annotations as ground truth and the mean Euclidian point-to-point distance as the error measure. The AAM and EAM methods were initialized by aligning the mean shape to a set of eye coordinates found by a Viola-Jones face/eye detection [24]. As the Viola-Jones failed to detect the eyes in nine images, the experiment were carried out on 110 images. For future comparison of this work it should be noted that Viola-Jones detected eye positions poorly in three out of the 110 images. This was mainly do to heavy reflections in glasses on the three persons in question. This has no consequence on the performed experiments as both AAM and EAM are initially aligned with the same eye coordinates.

2) *Face labeling with poor initialization*: The robustness of the AAM and EAM against poor pose initialization. In this experiment the AAM and EAM methods were initialized several time by regular perturbation of the Viola-Jones alignment (experiment 1) in intervals $\sqrt{2}[-20\text{pixel}, 20\text{pixel}]$, $[-15^\circ, 15^\circ]$ and $[90\%, 110\%]$ for diagonal translation, rotation and scaling, respectively. 189 (9 translations, 7 rotations, 3 scales) registrations were performed per leave-one-out step. The performance of the methods was evaluated similarly to experiment 1.

3) *Corpus Callosum segmentation*: The performance of AAM and EAM on medical images. This experiment used 62 two dimensional MR images (with different subjects) of the mid-sagittal cross-section of the corpus callosum brain structure recorded at the Danish Reserach Centre for Magnetic Resonance, Hvidovre Hospital. This data set is part of the

LADIS (Leukoaraiosis and DISability) study [17], a pan-European study involving 12 hospitals and more than 600 elderly subjects. Furthermore, each corpus callosum have manually been annotated with 72 landmarks by a clinician. The annotation were used as ground truth. The AAM and EAM methods were initialized several time by perturbation of the alignment between the mean shape and ground truth in the intervals $\sqrt{2}[-3pixel, 3pixel]$, $[-10^\circ, 10^\circ]$ and $[95\%, 105\%]$ for diagonal translation, rotation and scaling, respectively. The performance of the methods was evaluated with 6-fold cross-validation using the Dice, Jaccard and overlap coefficients to measure the quality of the segmentations. Each corpus callosum was segmented 75 times (5 translations, 5 rotations and 3 scales).

6 Experimental results

For all experiments principal component analysis was used for the matrix decomposition in Eqn. 1-3 and the truncation level was 0.95 of the variance of the training set. Lower and upper bounds of 2.5 times the standard deviation were enforced on the model parameters and α_1 was set to zero. Practical experiments showed $\alpha_2 = 0.001$ for the faces and $\alpha_2 = 0.001$ for the corpus callosum gave good results.

6.1 Face labeling with Viola-Jones initialization

Table 1 presents the results for the shape (annotation) displacement error. From the results it can be seen that EAM labels the face on average more than 10 per cent better than AAM. Visual examples of the face labeling are shown in Figure 3, where an initial red triangulation of the shape is shown as it progresses to an optimum. The green triangulation shape shows the hand annotated ground truth.

| Method | Displacement error in pixel | | | |
|--------|-----------------------------|-----------------|-----------------|-------------------|
| | Mean | 25th percentile | 75th percentile | 97.5th percentile |
| AAM | 4.8 | 3.1 | 4.7 | 15.2 |
| EAM | 4.3 | 3.1 | 4.5 | 8.8 |

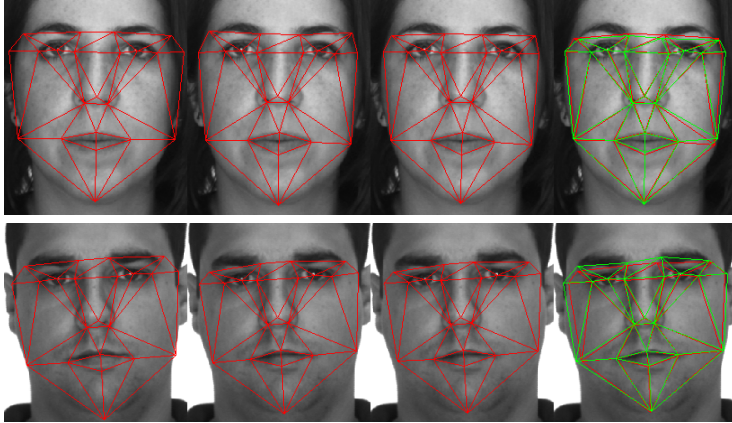
Table 1: Displacement error, initialized with Viola-Jones face/eye detection.

6.2 Face labeling with poor initialization

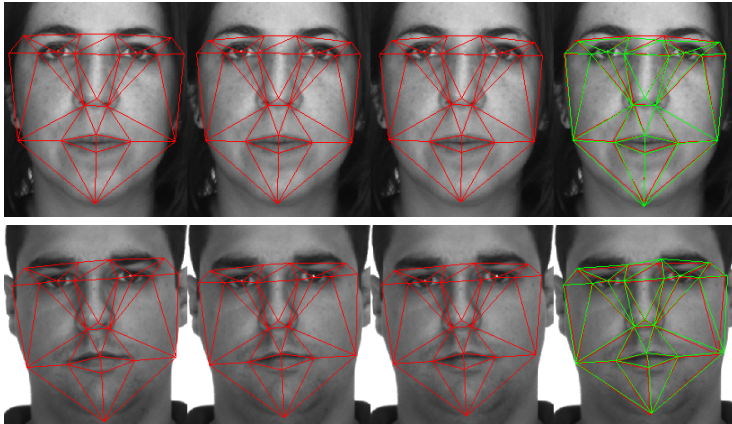
When making a poor initialization of the models it can be seen that EAM really outperforms AAM. By labeling the face on average 39 per cent better than AAM. Table 2 presents the displacement error results. Visual examples of the face labeling are shown in Figure 4.

| Method | Displacement error in pixel | | | |
|--------|-----------------------------|-----------------|-----------------|-------------------|
| | Mean | 25th percentile | 75th percentile | 97.5th percentile |
| AAM | 13.1 | 3.9 | 16.8 | 52.8 |
| EAM | 8.0 | 3.3 | 7.8 | 34.0 |

Table 2: Displacement error, initialized with poor face detection.

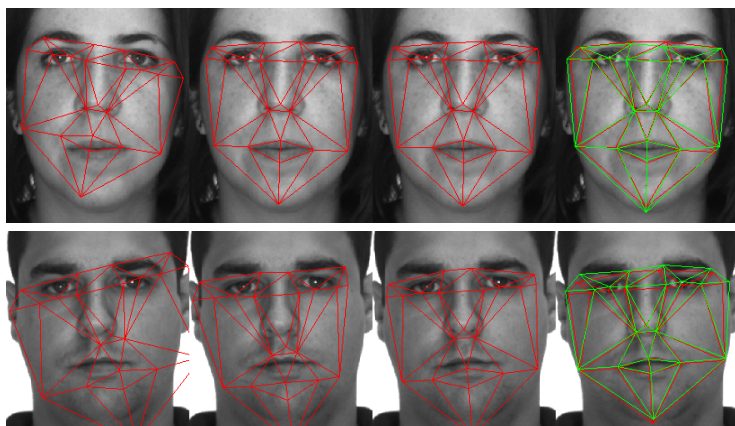


(a) AAM

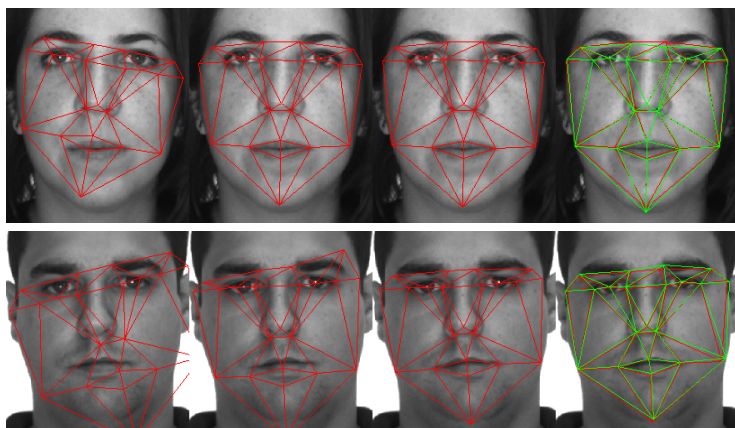


(b) EAM

Figure 3: Progress left-to-right from initialized mean shape with Viola-Jones face/eye detection to converged result for AAM and EAM, respectively. The red triangulation shows the shape as it progresses to an optimum. The green triangulation signifies the hand annotated ground truth.



(a) AAM



(b) EAM

Figure 4: Progress left-to-right from initialized mean shape with poor face detection to converged result for AAM and EAM, respectively. The red triangulation shows the shape as it progresses to an optimum. The green triangulation signifies the hand annotated ground truth.

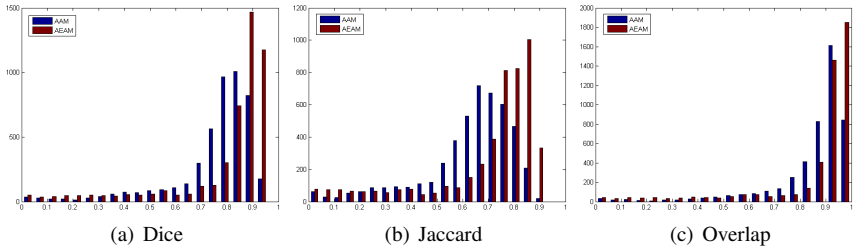


Figure 5: Histograms of the Dice, Jaccard and overlap coefficients for Corpus Callosum.

6.3 Corpus Callosum segmentation

The results of the experiment are listed in Table 3 and visualized in Figure 5. As the Dice and Jaccard coefficients after matching are 0.89 and 0.81 for AAM and 0.90 and 0.82 for EAM using only the unperturbed ground truth alignments for initialization the large differences in Dice and Jaccard coefficients primarily indicate that the EAM is more robust and secondarily that it is more accurate. This conclusion is supported by the histograms of the coefficients shown in Figure 5.

| Method | Mean | | |
|--------|------|---------|---------|
| | Dice | Jaccard | Overlap |
| AAM | 0.75 | 0.63 | 0.85 |
| EAM | 0.80 | 0.70 | 0.86 |

Table 3: Means for the Dice, Jaccard and overlap coefficients for Corpus Callosum segmentation.

7 Discussion and future work

The experimental results favour EAM against AAM. However, as the implementations of AAM and EAM are basic in the sense, that they do not incorporate any of the improvement mentioned in the previous work section, the difference in performance may be smaller as soon as the implementations are combined with some of the other improvements. Still, we believe that the experiments are good performance indicators.

The purpose of many of the AAM improvements is to speed-up the matching with real-time tracking being the end goal. It can be argued that we have gained robustness and accuracy with the EAM but lost efficiency as the problem in Eqn. 16 is much less cost efficient than e.g. the linear relationship strategy proposed by Cootes *et al.* [6]. However, if the application is real-time frame-by-frame tracking of an object the EAM formulation may not be as computationally expensive as it seems. First of all it is not necessary to do pose estimation as long as the object movement is slow compared to the framerate. Secondly, if you apply a BFGS optimizer you can use the estimate of the Hessian of the previous frame as initial estimate of cost function hessian which means that you may only need a small number of iterations to get to the optimum.

To train the shape model in EAM you need to estimate a small 3x3 strain covariance matrix for each triangle in the model (Eqn. 9) while you have to estimate a large covariance matrix and decompose it in PCA based AAM. A consequence of this is that fewer samples are needed to train the EAM than the AAM.

References

- [1] B. Abboud, F. Davoine, and M. Dang. Facial expression recognition and synthesis based on an appearance model. *Signal Processing: Image Communication*, 19(8):723–740, 2004. ISSN 0923-5965.
- [2] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004. ISSN 0920-5691.
- [3] R. Beichel, G. Gotschuli, E. Sorantin, F. W. Leberl, and M. Sonka. Diaphragm dome surface segmentation in CT data sets: A 3D active appearance model approach. *Proceedings of SPIE Medical Imaging 2002*, 4684:475, 2002.
- [4] C. M. Christoudias and T. Darrell. On modelling nonlinear shape-and-texture appearance manifolds. *Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1067–1074, 2005.
- [5] T. F. Cootes and C. J. Taylor. Constrained active appearance models. *Proceedings of the 8th International Conference on Computer Vision*, 2001.
- [6] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active Appearance Models. *Proceedings of the 5th European Conference on Computer Vision (LNCS)*, 1407:484–498, 1998.
- [7] G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using active appearance models. *Proceedings of the 5th European Conference on Computer Vision (LNCS)*, page 581, 1998.
- [8] X. Gao, Y. Su, X. Li, and D. Tao. A review of active appearance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 40(2):145–158, 2010. ISSN 1094-6977.
- [9] 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*, 10(1):19–40, 2006. ISSN 1361-8415.
- [10] R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(4):449–465, 2004. ISSN 0162-8828.
- [11] D. W. Hansen, J. P. Hansen, M. Nielsen, A. S. Johansen, and M. B. Stegman. Eye typing using Markov and active appearance models. *Proceedings of the sixth IEEE Workshop on Applications of Computer Vision*, pages 132–136, 2002.
- [12] F. Kahraman, M. B. Stegmann, S. Darkner, and R. Larsen. An Active Illumination and Appearance (AIA) Model for Face Alignment. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007*, pages 1–7, 2007.

- [13] R. Larsen, M. B. Stegmann, S. Darkner, S. Forchhammer, T. F. Cootes, and B. K. Ersboll. Texture enhanced appearance models. *Computer Vision and Image Understanding*, 106(1):20–30, 2007. ISSN 1077-3142.
- [14] D. Leimberg and M. Vester-Christensen. Eye tracking. Master’s thesis, Technical University of Denmark, 2005.
- [15] A. M. Martinez and R. Benavente. The ar face database. Technical Report 24, FGNet, 1998.
- [16] S. C. Mitchell, B. P. F. Lelieveldt, R. J. van der Geest, H. G. Bosch, J. H. C. Reiver, and M. Sonka. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. *IEEE Transactions on Medical Imaging*, 20(5):415–423, 2001. ISSN 0278-0062.
- [17] L. Pantoni, A. M. Basile, G. Pracucci, K. Asplund, J. Bogousslavsky, H. Chabriat, T. Erkinjuntti, F. Fazekas, J. M. Ferro, M. G. Hennerici, J. O’Brien, P. Scheltens, M. C. Visser, L. O. Wahlund, G. Waldemar, A. Wallin, and D. Inzitari. Impact of age-related cerebral white matter changes on the transition to disability—the LADIS study: rationale, design and methodology. *Neuroepidemiology*, 24(1-2):51–62, 2004. ISSN 0251-5350.
- [18] X. Pennec, R. Stefanescu, V. Arsigny, P. Fillard, and N. Ayache. Riemannian Elasticity: A Statistical Regularization Framework for Non-linear Registration. *Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention (LNCS)*, 3750:943–950, 2005.
- [19] S. M. Pizer, P. T. Fletcher, Y. Fridman, D. D. Fritsch, A. G. Gash, J. M. Glotzer, S. Joshi, A. Thall, G. Tracton, P. Yushkevich, and E. L. Chaney. Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision*, 55(2):85–106, 2003. ISSN 0920-5691.
- [20] Mark Schmidt. minfunc and minconf optimization package, 2009. <http://www.cs.ubc.ca/~schmidtm/Software/minConf.html>.
- [21] I. M. Scott, T. Cootes, and C. Taylor. Improving appearance model matching using local image structure. *Proceedings of Information Processing in Medical Imaging*, 2003, pages 258–269, 2003.
- [22] M. B. Stegmann. Object tracking using active appearance models. *Proceedings of the 10th Danish Conference on Pattern Recognition and Image Analysis*, 1:54–60, 2001.
- [23] J. Sung, T. Kanade, and D. Kim. A unified gradient-based approach for combining ASM into AAM. *International Journal of Computer Vision*, 75(2):297–309, 2007. ISSN 0920-5691.
- [24] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2002.