Hands-on Experience with Active Appearance Models

Hans Henrik Thodberg[1], Technical University of Denmark, Dept of Informatics and Mathematical Modeling

ABSTRACT

The aim of this work is to explore the performance of Active Appearance Models (AAMs) in reconstruction and interpretation of bones in hand radiographs. AAM is a generative approach that unifies image segmentation and image understanding. Initial locations for the AAM search are generated by an exhaustive filtering method. A series of AAMs for smaller groups of bones are used. It is found that AAM successful reconstructs 99% of metacarpals, proximal and medial phalanges and the distal 3 cm of radius and ulna. The rms accuracy is better than 240 microns (point-to-curve). The generative property is used (1) to define a measure of fit that allows the models to self-evaluate and chose between the multiple found solutions, (2) to overcome obstacles in the image in the form of rings by predicting the missing part, and (3) to detect anomalies, e.g. rheumatoid arthritis. The shape scores are used as a biometrics to check the identity of patients in a longitudinal study. The conclusion is that AAM provides a highly efficient and unified framework for various tasks in diagnosis and assessment of bone related disorders.

Keywords: Active appearance model, shape models, segmentation, image understanding, arthritis, osteoporosis

1. INTRODUCTION

Active Appearance Models (AAMs) were introduced recently by Cootes and collaborators as a new efficient method for computer vision and medical image analysis [1,2]. AAM does not fit into the conventional classification of medical image analysis, because it incorporates elements of segmentation, registration, recognition and image understanding. The underlying idea is that these elements of vision are not to be dealt with in separation, but rather in a unified framework: It is sometimes necessary to recognize and understand the image in parallel with the segmentation of the objects in the image. The AAM contains a huge amount of prior knowledge of the objects to be found, which constraints the search for the object and as a result the segmentation is much faster and more robust.

AAM can be seen as an example of machine learning, analogous to neural networks in many ways: AAM makes use of a large amount of memory to store knowledge gained using a training set. Furthermore AAM uses distributed representation and processing, there is feedback between two levels of cognition (segmentation and understanding) and AAM is non-rule-based and hence the behaviour cannot be explained in sequential language; rather it appears intuitive.

The adaptation of AAM is made possible by the large memory available on today's computers. Computation time is not an issue, in fact AAM is less CPU intensive than the deformable template models used ten year ago, which were typically based on simulated annealing. Memory requirements are even today an obstacle to the extension to 3D. This paper deals only with 2D images: high resolution radiographs (ordinary X-ray images) of the hand in PA projection, which is the default modality used to assess rheumatoid arthritis and DXR-bone mineral density [3].

The motivation for this work is to form the basis for Computer-Aided Diagnosis on hand images and similar images.

2. METHOD

The reader is assumed to be familiar with the Active Appearance Model, e.g. from the review by Cootes and Taylor [2]. The section describes the details of the implementation.

2.1 Data and Annotation

To define an AAM, an annotated training set is required. This was obtained by drawing the contours of the hand bones on 100 images of women spread evenly in the age range 20-80 years and from four continents (Europe, the UK, North America, Argentina, Korea and China). For the metacarpals 100 additionally examples were annotated yielding in total of 200 examples. All data were collected as part of the clinical documentation for the X-posure System, which uses the metacarpals for the DXR-BMD measurement [3]. An example of an image is shown in figure 1.

---

[1] hht@imm.dtu.dk, address: IMM, DTU, 2800 Lyngby, Denmark

Figure 1: A standard hand radiograph digitised in 300 dpi on a UMAX Powerlook III scanner

2.2 Definition of marks

For metacarpals, the bone axis is determined and relative to this a number of marks are defined to represent the shape, see fig 2.
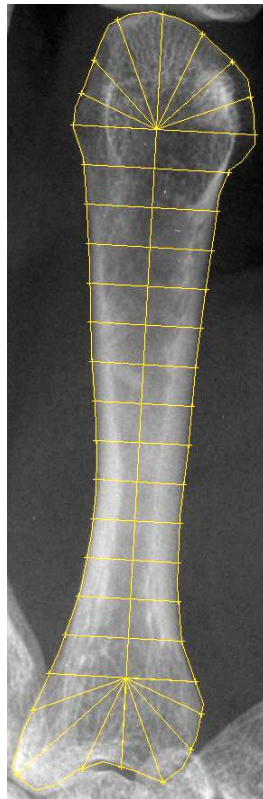

Figure 2: An annotated metacarpal 2 shown with the definition of marks. The bone axis is found and marks are defined at regular intervals relative the axis. In the ends, lines are shown radiating at regular angles.

In the bone ends, the marks can be defined at constant angle intervals as in figure 2 - However, when the shape is far from a half circle, as in the proximal end of the metacarpal, the marks are defined instead at constant intervals of arch length.

2.3 AAM essentials

AAM is based on a principal component analysis (PCA) of shape variation together with a PCA of the image intensity variation. The latter is computed after shape-alignment and triangulation anchored in the shape marks, see figure 3. The two PCA models are able to generate quite faithful images. This generative nature of AAM is a central theme in this work.
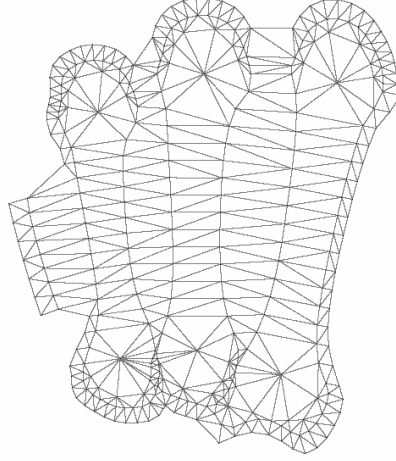

Figure 3: The triangulation spanning the template used by AAM to locate metacarpals 2-3-4. Part of metacarpal 5 is included to avoid confusion between metacarpals 2-3-4 and 3-4-5.

In standard AAM the shape scores and the image intensity scores are processed using a further PCA to obtain the combined scores denoted c. The four Euclidian parameters defining the location, orientation and scale of the template are called the **pose** and denoted t. The combined scores and the pose are used as primary dynamic variables and are driven by a residual image defined as the difference between the observed image intensities and the intensities generated from the current c-scores:

$$\delta g = g - g_{gen}$$

$$\delta c = B_c \, \delta g$$

$$\delta t = B_t \, \delta g$$

This work uses a simpler and faster version of AAM, where only the shape scores b and the pose t are used as dynamical variables.

$$\delta g = g - g_{proj}$$

$$\delta b = B_b \, \delta g$$

$$\delta t = B_t \, \delta g$$

Instead of using the generated image g_{gen} as in standard AAM, the observed intensities g are simply projected onto the PCA intensity model (defined by the mean image g_{mean} and the eigenvectors matrix P_g):

$$g_{proj} = g_{mean} + P_g \, (P_g{}^T \, (g - g_{mean}))$$

This method is described in section 9.2 of [2] and is useful in situations, where the dimension of the intensity PCA model is considerably larger (here 30) than the dimension of the shape model (here 12). It also leads to a speed-up of the AAM algorithm (see [2]).

2.4 The full AAM recognition system

A single AAM search described above cannot solve the complete reconstruction problem in an application. A series of AAM searches must be used, and the initialisation and quality measure must be dealt with.

Initial locations for the AAM search are generated by an exhaustive search by a convolution filter for the occurrence of three aligned bones with a certain separation (see [3] for details), providing of the order of 10 candidates for the location of metacarpal 2-3-4. All candidates are used as initial locations for the AAM search for metacarpal 2-3-4 in 41 dpi, followed by another AAM working in 75 dpi. The search resulting in the best quality fit is selected (see below on quality measure).

Based on the location of the three metacarpals, an AAM for the second proximal phalanx is involved. This is initialised by predicting the proximal and distal ends of the bone based on the (t and b) found for the metacarpals. The distal end cannot be accurately predicted so four extra initialisation locations around the predicted are also used. An AAM is initiated with each of these and the quality measure selects the correct one. This principle is repeated with separate AAM models for the two other proximal phalanges and for the medial phalanges. Finally one AAM for the distal 3 cm of radius of ulna is invoked.

The resolution of the intensity models is 75 dpi (0.340 mm pixel size), and only one resolution is used – except for the metacarpal model which is larger – here a 0.7 mm pixel model is used for the first part of the search to save time.

2.5 The quality measure

Several different initialisations are typically used to find each structure, and it is important to design a quality measure on the resulting convergence which is able to select the correct one, i.e. it is used comparatively. In addition it also must be able to qualify the solution on an absolute scale and determine whether an object of the desired type was found. AAM is in a good position to do this because it is very specific, but this topic has found no completely satisfactory solution yet. In the present work the following quality measure is used:

$$\text{misfit} = \tfrac{1}{2} \left(\text{mean}((g - g_{\text{proj}})^2) + \chi^2_{\text{Tex}} / \text{ndf} / 10 \right)$$

The first term is the average square deviation between the observed and the generated image. The second is the sum of the squared image scores, divided by ndf=30 to bring its expected value close to 1, and divided by 10 to bring it onto the same scale as the first term.

2.6 Blind spots

A common problem in X-ray images is the presence of metal objects, for example rings on the phalanges, or hip implants. It is desirable to be able to locate the anatomical structure automatically even in the presence of such objects.

Areas containing metal objects can easily be detected in X-ray images, because steel and heavier elements absorb X-rays far better than bone and soft tissue, so these areas are whiter than any anatomical structure on a well-exposed X-ray image.

The solution to the problem presented here therefore first marks these areas as *blind spots*. The idea in the solution is that the image inside the blind spots is restored *in parallel* with the convergence of the AAM. Initially, the intensity values in a blind spot are set to the mean computed at the border of the spot.

As usual, a suitable initial guess of the location of the shape is given from another part of the system. Then the first iteration of convergence is performed. When the pose t and shape b has been updated and g has been sampled, the projected image g_{proj} is used to update the blind spots. This cycle is continued until convergence as illustrated in figure 8.

3.   RESULTS

3.1 Reconstruction Error

The implemented AAMs successfully reconstruct metacarpals 2-4, proximal and medial phalanges 2-4 and the distal 3 cm of radius and ulna. Examples are shown in figure 4-6.

The performance is evaluated by the rms (root mean square) *point to curve distance*, p2c: For each mark the shortest distance to the boundary defined *by the manual annotation* is computed. For metacarpals the p2c is computed excluding the proximal ends, since these are hard to segment even for the human reader.

Table 1 Rms point to curve (p2c) errors

	Number of Observations	p2c (microns)	p2c (microns) 4 fold cross validated	p2c distal end (microns)	p2c distal end (microns) 4-fold cross validated
Metacarpal 2-3-4 (excl proximal part)	200	340		210	
Proximal phalanges	100	180		130	
Medial phalanges	90	130	180	90	120

In these results the training data were also used for validation, which will always lead to better performance than using new data. To estimate this effect, one of the models, the medial phalanx model, was subjected to 4-fold cross validation of the entire model estimation process. The errors increase by approx 35%. The performance is therefore summarised as **cross validated p2c<240 microns on proximal and medial phalanges**.

Notice that the manual annotation is not identical to the PCA model representation of the contour. In fact the p2c distance between the manual and the PCA representation is of the order of 50% of the numbers in table 1.

Figure 4: The AAM reconstruction of metacarpals 2-3-4. Some reconstruction error is visible for instance in the shaft of metacarpal 3. Also shown are the results of three other AAMs reconstructing phalanges 2, 3 and 4.
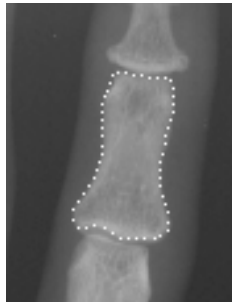

Figure 5: AAM reconstruction of a medial phalanx.

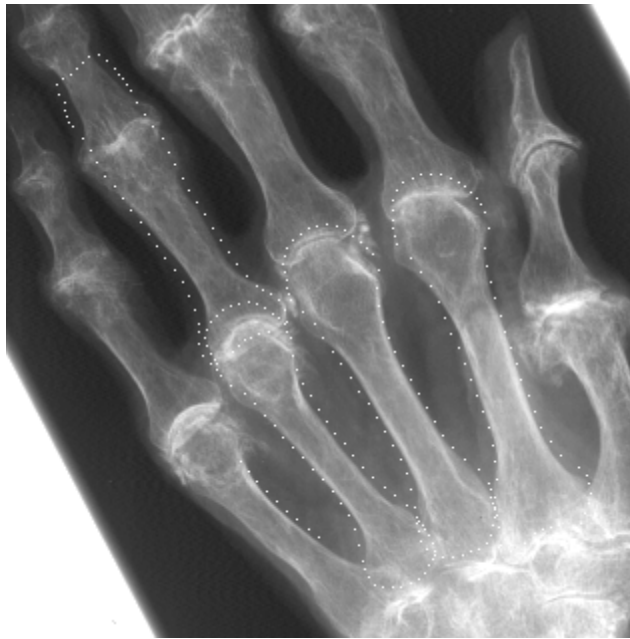**Figure 6:** AAM reconstruction of the distal ends of radius and ulna

Figure 7: An example of a reconstruction with a large misfit due to rheumatoid arthritis.

The misfit parameter defined in section 2 is typically between 0.05 and 0.10. A cut-off is defined to 0.18 for the misfit measure. If the misfit is better than this cut-off, the reconstruction is always correct - this has been validated on 400 independent images. If the misfit error is larger than the cut-off, a warning is issued. This occurs in 2-4% of the cases and means that the reconstruction can potentially be wrong and the user must therefore check the result. Around 1% of all cases yield wrong reconstructions - again validated on 400 independent cases.

Among the cases with a warning are cases like figure 7 showing diseases characterised as a deviation from normal morphology. The warning is hence very useful in the practical work with these images in research and clinical practice, because it addresses the attention to images with errors or anomalies.

3.2 Blind spots

The blind spot method is tested in an AAM that locates metacarpal 2-3-4, in this case it is a special model that omits the proximal ends from the model in order to focus on getting the distal ends right. As shown in figure 8a, a rather large blind spot is defined in order to challenge the model. Figure 8b shows the result of the first iteration. The image in the blind spot is more correct, and the convergence is therefore already much less disturbed. In fact, the texture of the blind spot is aligned to the current model of the shape and pose, so that the texture inside the blind spot is neutral towards driving the shape in one direction or another. This means that the method works for even very large blind spots. Here the spot is approx. 25% of the template area.

The blind spot method illustrates the power of the generative model very clearly. The reconstructed area is virtually indistinguishable from the surroundings, which shows that even the rather small 30-dimensional image model and 12-dimensional shape model is able to understand the anatomy remarkably well.

With this solution, rings on the phalanges are no longer a problem for AAM, as long as they are not made of light materials like wood, aluminium or ivory. The method will also allow the reconstruction of the femur in the presence of various steel implants.

Finally, the method can be extended to not require the blind spots to be defined in advance. Instead the model would detect deviating spots during convergence based of large values of δg in regions of a certain size. This could be used to detect ivory or bone rings, or to detect local anomalies like bone metastases.

Figure 8: (a) The blind spot and the initial location of the model shape. (b) Result of first iteration. (c) Result of the sixth iteration. (d) Result of iteration 29 - the blind spot is invisible.

3.3 Biometrics

The shape scores determined by the AAM reconstruction can be used as a biometrics to identify subjects at a later X-ray visit. The bone shapes change only very slowly in healthy adults, since bone loss occurs mainly at the endosteal surface and within trabeculae. Also the size parameter - one of the four pose parameters - is expected to be very constant. It is determined independently from the shape scores in AAM, so the size can check for magnification effects, or it can be used as an additional biometrics in the identification.

This feature of AAM was tested on images from 11 subjects followed over 3 visits with 18 month between the first and the last visit. Three images were taken at each visit with repositioning of the hand, yielding in total of 9 images per

subject. A given biometrics (a shape score or the size) is denoted m, and the determinations from the nine images are indexed m_{rv} with the repetition number r and the visit number v.

Short-term precision is computed across repetitions in the same visit as follows. For a single subject and visit it can be estimated as st.dev(m_{11}, m_{21}, m_{31}), where st.dev uses division by sqrt(3-1). The precision is computed as the rms of the 3·11 different estimates that can be made.

Likewise, long-term precision is computed across visits, with fixed repetition number. One such estimate is st.dev(m_{11}, m_{12}, m_{13}), and the precision is computed as an rms of 33 such estimates.

The population spread σ_{pop} is also computed from the 11 subjects, and the coefficient of variation cv is formed as the ratio

$$cv = precision / \sigma_{pop}$$

since this is the relevant quantity in biometrics.

For the metacarpal 2-3-4 model the short and long term cv's are shown in table 2. The short term cv is due mainly to the slightly different positioning of the hand in the three exposures.

The long-term cv includes in addition to the short term effects also effects from biological changes in the subjects, aging of X-ray tube, change of developer, change of film batch etc. Also a different scanner was used for scanning visit 3 (UMAX rather than Linotype). The long term precision includes all the relevant effects to consider when evaluating these quantities as a biometrics in practice. The size and the first four shape scores have precision of the order 30% of the population spread, while the higher scores give approx 50%.

Table 2. Precision of AAM shape biometrics Table of short- and long-term precision of AAM determination of shape size and scores for metacarpals

	Short term cv	Long term cv
Size	0.23	0.30
score 1	0.20	0.27
score 2	0.20	0.26
score 3	0.22	0.35
score 4	0.15	0.23
score 5	0.31	0.53
score 6	0.40	0.55
score 7	0.34	0.56
score 8	0.39	0.57
score 9	0.41	0.48
score 10	0.32	0.39
score 11	0.40	0.60
score 12	0.63	0.97

The size and the twelve shape scores are to a good approximation uncorrelated and independent, so the population spread becomes the product of the spreads. When comparing a baseline quantity with one measured at a follow-up, cv$\sqrt{2}$ is the error relative to the population spread, i.e. the posterior/prior volume ratio. To posterior/prior volume ratio with 13 quantities is therefore

$$V_{post}/V_{prior} = \sqrt{2}\ cv(size)\ \sqrt{2}\ cv\ (shape1) \ldots \sqrt{2}\ cv(shape\ 12)$$

This volume ratio is shown in table 3 for metacarpals 2-3-4 as well as for the AAMs for phalanges 2, 3 and 4.

Table 3: Precision of AAM parameters as biometrics.

AAM model	V_{post}/V_{prior}
Metacarpals 2-3-4	0.0016
Phalanx 2	0.0015
Phalanx 3	0.0008
Phalanx 4	0.0008

The phalanges models tend to have better precision than the metacarpal model, presumably because the proximal ends of the metacarpals are difficult to locate.

It is concluded that from a single AAM model, the shape determination is able to identify subjects with a specificity, which in rough terms are given as

$$\text{specificity} = 1 - V_{post}/V_{prior} = 0.999$$

This means (roughly) that only 1 in 1000 random images are as close to the baseline as the correct follow-up image.

Biometrics is useful in cohort studies, where it is essential that the identity of the subjects is controlled. This feature can also be used to reveal if a subject is erroneously included twice in a study. Finally it may be helpful to correct errors in ordinary clinical practice.

4. CONCLUSION

AAM provides a highly efficient and unified framework for various tasks in diagnosis and assessment of bone related disorders. If forms the bases for several computer-aided radiology applications, e.g. in rheumatoid arthritis assessment. The simplified version of AAM which uses shape scores and pose as dynamic variable works very well. The accuracy is better for phalanges than for metacarpals, which are harder to segment at their proximal ends. The cross-validated accuracy of the phalanges model is better than 240 micros. The blind spot method solves the problem with rings and other clutter in the images. The shape of the bones is determined with a precision that makes it useful for checking the identity of serial X-rays measurements; again the performance is better on phalanges than on metacarpals.

REFERENCES

1. T. F. Cootes, G. J. Edwards and C. J. Taylor: "Active Appearance Models" in *Proc. European Conference on Computer Vision* 1998, H. Burkhardt and B Neumann Ed.) Vol. 2 Springer 1998

2. T. F. Cootes, C. J. Taylor: "A detailed report about Active Shape Models and Active Appearance Models", on http://www.isbe.man.ac.uk/~bim/ (Postscript), (2001).

3. H. H. Thodberg and A. Rosholm, "Application of the Active Shape Model in a commercial medical device for bone densitometry", in *Proceedings of British Machine Vision Conference, 2001*, ed. T.F. Cootes and C.J. Taylor, p 43-52, published by Digital Repro, Stockport , UK