Application of the Active Shape Model in a commercial medical device for bone densitometry

H. H. Thodberg and A. Rosholm,
Pronosco, Kohavevej 5, 2950 Vedbaek, Denmark
hht@pronosco.com

Abstract

Osteoporosis is a common disorder characterised mainly by low bone mineral density (BMD), and leading to an increased risk of fracture. We have developed a new device that estimates BMD from ordinary hand radiographs. A crucial element of this method is the reconstruction of the metacarpals. This paper describes how this was solved using the Active Shape Model (ASM). Standard ASM is unable to locate the metacarpal shafts due to the lack of dynamics in the direction along the bones. Therefore ASM was extended with a translation operator, which solved the problem. A hierarchical filtering method was used to construct a sufficient list of initial guesses for the ASM. The performance of ASM and the experience with the integration of ASM in a commercial medical device is reported. The ASM achieves 99.5% reconstruction success and is able to validate its own reconstruction in 97% of the cases. The system (Pronosco X-posure system, version 2.0) has been approved by the FDA, and more than 100 units have been sold.

The concept of the translation operator is generalised to the More Active Shape Model (MASM), which also allows a natural integration with the Active Appearance Model.

1 Introduction

Osteoporosis is a condition of reduced bone mass (and to some extent also poor bone architecture), which affects the whole skeleton to a similar degree and leads to a reduced strength of the bone and therefore an increased risk of fractures, in particular at the hip, the vertebrae and the wrist. Approx. 20% of all women suffer an osteoporotic hip fracture in their lifetime, while men are affected at a 5-10 percent level. Osteoporosis can to a large extent be prevented. For instance, the frequency of hip fracture is reduced by typically 50% by medication. The diagnosis requires measurement of the bone mineral density (BMD) - the weight of calcium per projected bone area - but the percentage of people diagnosed is very inadequate. The most common diagnostic method today is dual energy X-ray absorptiometry (DEXA), which is performed on large dedicated scanners. There is a need for alternative, more accessible methods of BMD measurement, and we have therefore developed a new method, Digital X-ray Radiogrammetry (DXR) that estimates a BMD value from an ordinary radiograph of the hand. The system determines DXR-BMD in the shafts or the three middle metacarpals as indicated in figure 1. In these regions the bone has the structure of a cylindrical tube and

BMVC 2001 doi:10.5244/C.15.6

the width W and cortical thickness t can be determined with very high precision and accuracy. BMD is then estimated from the formula

$$\text{DXR-BMD} = c\ \pi\ t\ (1 - t/W)$$

where c can be interpreted as the mineral mass per volume of compact bone.

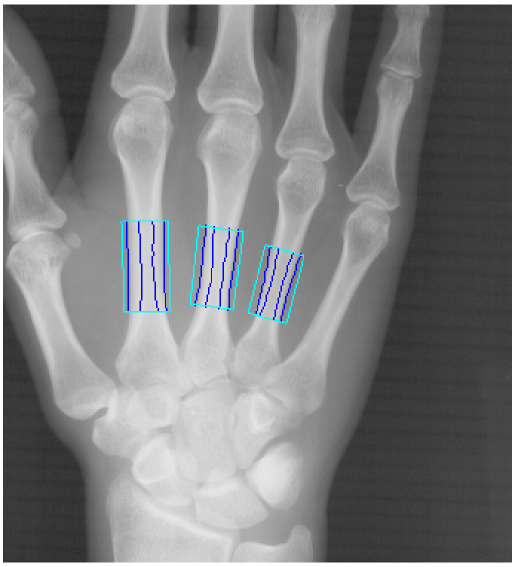

Figure 1: A standard hand image, and the regions of BMD determination

2 The machine vision problem

This paper addresses the machine vision problem of reconstructing the three shafts from which the DXR-BMD measurement is performed. This appears at first to be a rather easy task but the system must be able to work robustly on an extremely wide range of biological variations and technical conditions. Figure 2-4 show examples of images where the ends of the bones appear differently, where the exposure is extreme or where the placement of the hand is unusual.

In general the reconstruction must work irrespective of the appearance of the image outside the sub-region containing the three shafts. Thus, it cannot be assumed that metacarpal 1 or 5 or the phalanges are in the image - the recognition must be based on the appearance of the shafts themselves. No particular exposure level, contrast or particular orientation of the hand can be assumed, and the system must determine whether it is a left or a right hand.

The resolution of the image is given; usually it is 300 pixels per inch.

The reconstruction must have excellent *sensitivity*, i.e. it must detect the three metacarpals if they are present.

The reconstruction must also have excellent *specificity*, i.e. it should reconstruct only metacarpals 2,3,4 and <u>not</u> reconstruct other groups objects that have some similarity with metacarpals 2,3,4.

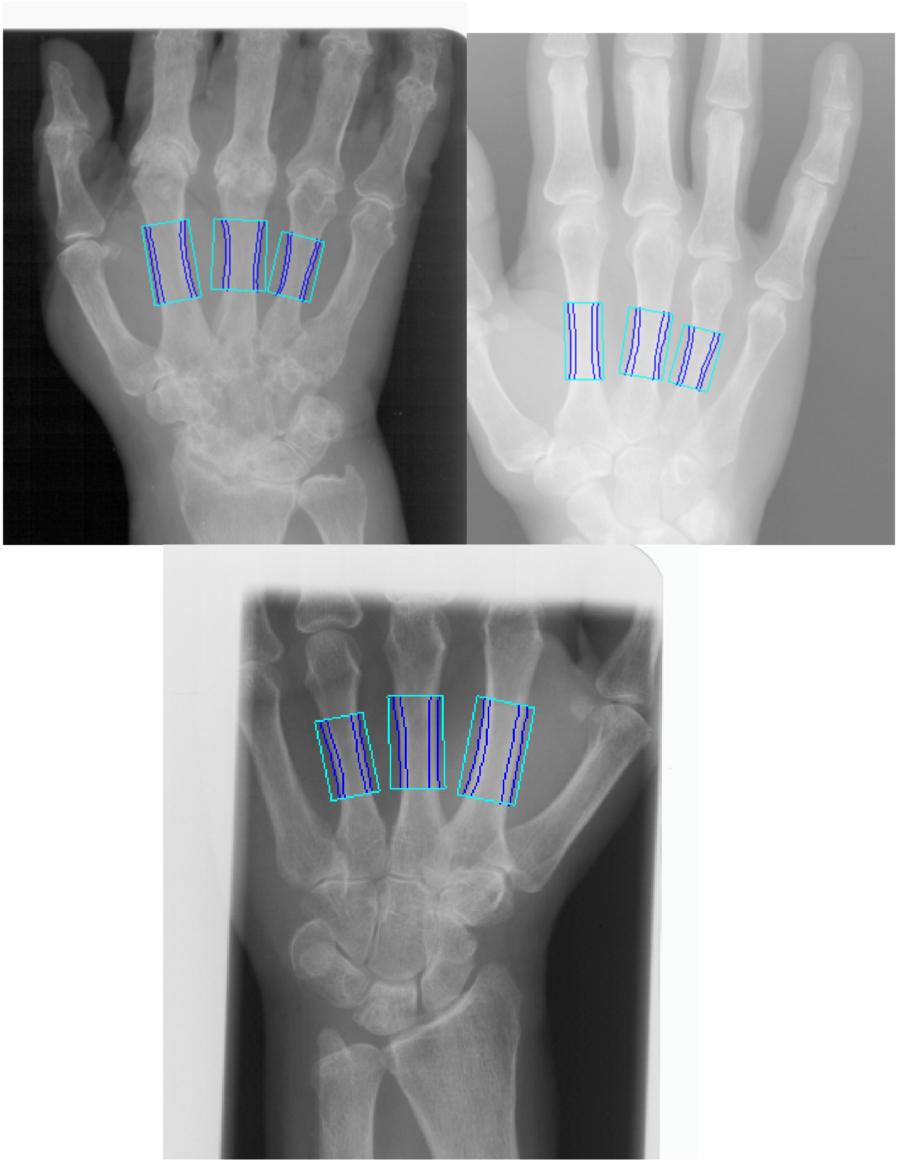

Figure 2-4: Illustrations of the variability of the images to be reconstructed

3 The method

3.1 Selection of model - the Active Shape Model

Humans receive a lot of guidance in reconstructing by viewing the surroundings, and smoothly switching between using whatever extra information is present in the images. This is a very complex strategy, which is hard to automate, why the machine vision solution instead uses only the information contained in the shafts themselves. This requires a much more limited knowledge and experience of the anatomy, which is easier to implement.

The requirements of high sensitivity and high specificity led to adopt the active shape model combined with an exhaustive search for initial locations for the ASM. ASM is well known to be specific and the exhaustive search complements ASM by giving high sensitivity.

The active shape model was formulated by Cootes and his collaborators in 1992 [1]. Its general application for medical images was presented in 1994 [2], and Edwards applied ASM to reconstruct hand X-rays of children already in 1994 [3]. Since then, the adaptation of ASM in medical applications has been relatively slow. The most serious application published so far is in reconstruction of vertebrae for detailed diagnosis of osteoporosis [4]. There is no reason for this slow adoption by applications: ASM is easy to understand and implement and it does not require a lot of CPU or RAM.

In 1998 Cootes and collaborators launched AAM [5], which is more powerful, more complicated, and more computer-intensive. Hopefully this will not divert interest from ASM – both ASM and AAM will have a role to play in the future. ASM recognises the objects based on edge signatures in the image, while AAM matches the complete intensity distributions. Sometimes edges are the most important, at other times the intensity patterns are important – and sometimes both are important. Section 5 presents an improved version of ASM, called the More Active Shape Model or MASM. It is also guided by edges but uses elements of AAM. This model is a step towards merging ASM and AAM to a more complete model.

3.2 The ASM machinery

The primary dynamical variables in ASM are a set of <u>marks</u> (points) defined on boundaries or landmarks of the object. The marks define <u>the shape</u> to be reconstructed.

Annotation

To construct the model, these marks must be placed manually on a set of training examples. Although this annotation task may seem tedious, it is in fact an efficient way to acquire knowledge from an expert.

Shape alignment and PCA

The shapes of the training set are aligned using Procrustes analysis, and a principal component analysis (PCA) is used to describe the variation of the aligned shapes among individuals.

Initialisation

New images are reconstructed in the following manner. The algorithm is initialised by placing the average shape approximately correct on the image.

Moving shape to edges

The image is now sampled at each mark on a linepiece perpendicular to the shape boundary. This sample is searched for a match to an edge model. The match is defined either by maximum gradient, or by minimum distance to an edge signature based on a Mahalanobis measure derived from the training set specifically for each mark. The latter is the most efficient, but requires more computation. The marks are moved to the found edges, thus generating a new shape.

Regularising the shape

The best fit of the marks to the shape model is now found (this requires an iterative procedure). The shape model has four Euclidean (or pose) parameters and a number of principal component shape scores. The scores are not allowed to exceed 3 standard deviations. The shape is now replaced with this regularised shape.

Thus reconstruction proceeds in this way by alternating between moving the shape to the edges and regularising the shape until convergence.

Details of the model

The actual implementation of ASM uses 103 training examples of women age 20-80 from USA, China and Denmark. It works in a 75 dpi resolution, and uses 8 principal components.

The noise of the one-dimensional samplings perpendicular to the marks is reduced by utilising a five pixel wide band rather than a single-pixel wide line. The first iterations use maximum gradient to define the edge while the later iterations use Mahalanobis distance to an edge signature defined by a 11-window containing the gradients normalised by the sum of their absolute values, as in standard ASM. The extent of the line search is reduced in the later part of the convergence.

3.3 Initialisation of ASM for metacarpal shafts

ASM is not a complete recognition method, because the search has a limited basin of attraction around the correct shape, and it must be initialised within this. There is no general method that can provide a list of initialisations that is both certain to hit the basin and at the same time reasonably short.

A method was developed for the problem at hand. It relies on the existence of characteristic elements or signatures that can be found by simple filtering.

The entire image is searched for occurrences of <u>shafts</u>. This is done by filtering the image in a reduced resolution with a small template consisting of a white strip in a black square in 4 different rotations with subsequent finetuning of the direction search near the maximum of the four. At each pixel is recorded the largest correlation to a shaft template as well as its angle. These two quantities are combined in figure 5 to a colour image of the strength and orientation of shafts in the image.

It is conceivable that the brain uses similar preprocessing to form useful representations of primitives, which are used for various higher-level processing tasks. For locating faces, Cootes have used the signature of a pair of eyes for a similar initialisation procedure [Tim Cootes, private communication] and the principle may work for a wide class for problems.

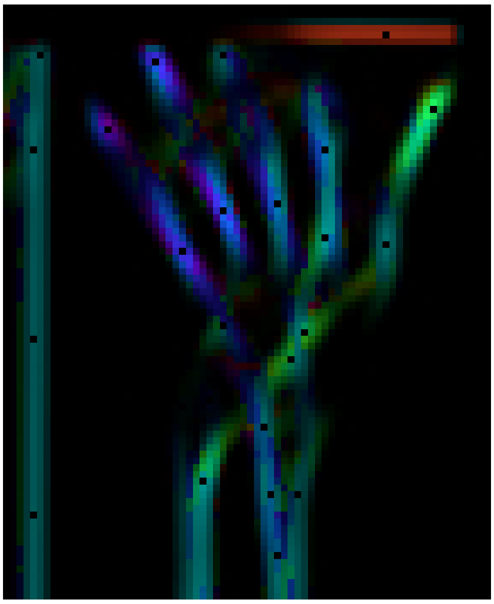

Figure 5: The filtering that locates bone shafts

A further search is made in the vicinity of local maxima in this image for the occurrence of three bones with the proper orientations. This results in a list of guesses, on average 8 per image. This method is exhaustive, but since filtering is fast (using Intel's image processing library) it takes less than two seconds on a 500 MHz PC. Among 5000 cases the method never misses the correct location of the three metacarpals, i.e. the sensitivity is 100%.

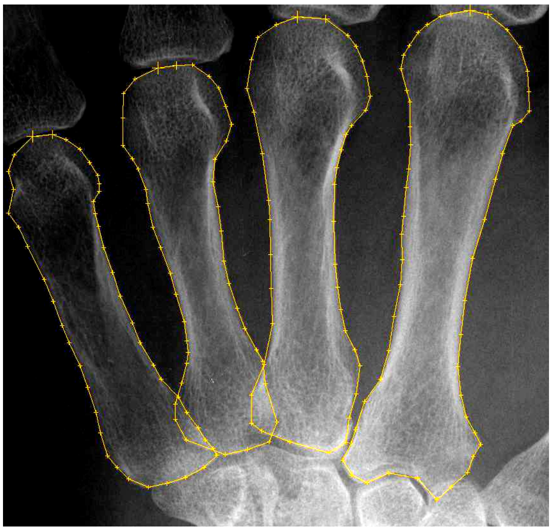

Figure 6: The manual annotation

3.4 Annotation

Metacarpals have no natural landmarks, so to define marks, the entire metacarpal boundary is annotated. The metacarpals are treated one by one:

- The contour is annotated by placing points at rather arbitrary intervals in the original 300 dpi resolution, illustrated in figure 6.
- The bone axis is defined and bone coordinates are measured along this axis with coordinate 0% at the proximal end and 100% at the distal end. The axis is defined as the middle of the bone at coordinates 30% and 70%. This is an iterative algorithm.
- In the shaft region the marks are placed at evenly spaced bone co-ordinate values, e.g. at 15, 20, 25,…, 85%. In the proximal/distal ends the marks are placed at the intersection of lines radiating from the 15/85% point on the axis at evenly spaced angles, see figure 7. In the end, only marks on the shaft were used, in total 42 marks shown in figure 8.

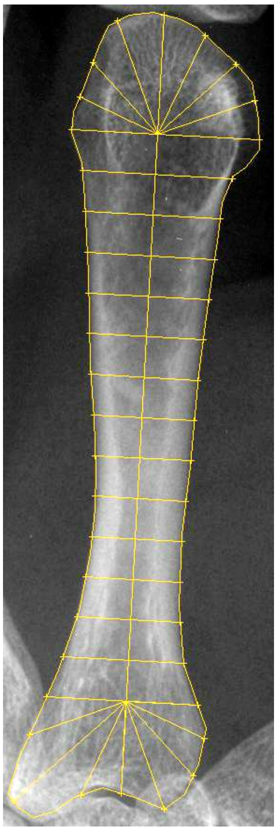

Figure 7: The algorithm used to transform the annotation points of figure 6 into the marks for ASM models is.

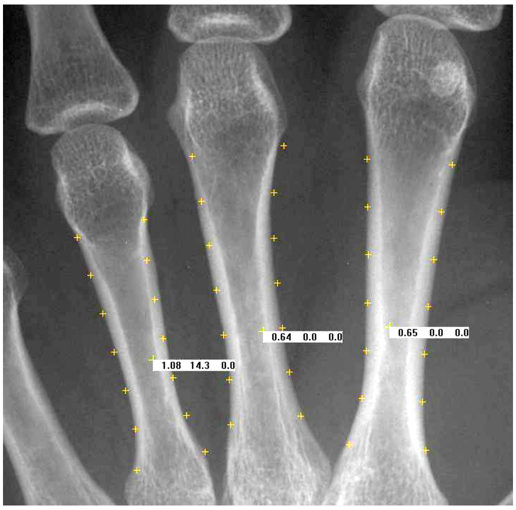

Figure 8: The 42 marks used in the ASM, here shown as reconstructed by the ASM.

3.5 The translation operator

The standard ASM cannot easily reconstruct a shape described by the 42 marks in figure 8. ASM moves the model-shape by moving the marks perpendicular to the contour. Here, the marks are placed only on the left and right of the object, i.e. there are no marks in the top and bottom of the object. Therefore, there are no forces to drag the marks up and down along the bone. ASM has in fact mainly been used for rather compact objects like faces where shifts in all direction can easily be obtained.

Inspired by AAM, a translation operator was introduced in the following way: A shaft is characterised by having a minimum width in the middle rather than towards one of the ends. When the 14 marks of a shaft are sitting too low on the image shaft, there is a characteristic deviation between the model shape, which has the minimum in the middle, and the shape inferred by the image edges, which have the minimum off-set from the centre. The difference of these two patterns of bone widths is proportional to the translation that will bring the marks into the proper position. Hence this difference is computed and used to translate the marks. This gives an extremely robust algorithm that very quickly homes in on the middle of the shafts.

3.6 The quality measures

The specificity of ASM allows it to self-evaluate its reconstruction of the anatomy. A misfit measure is constructed as the sum of two terms

$$\text{Misfit} = \text{Deformation} + \text{EdgeErrors}$$

“Deformation” is the chi-square / 8 of the standardised shape scores, which has mean 1 on the training set. The scores of the shape model are explicitly truncated during reconstruction so each term in the chi-square is at most 9 – hence the “Deformation” is always less than 9.

EdgeErrors is the percentage of marks for which the image edge is more than 0.8 mm away from the model shape.

The misfit measure is used for two purposes.

Firstly the ASM is initiated and run to convergence on each of the on average eight initial guesses given by the filtering method described in section 3.3. The ASM chooses the solution with smallest misfit.

Secondly this minimum misfit is used as an absolute measure of confidence in the reconstruction.

The misfit is typically less than 5. If it is larger than 30, the recognition reports no found object. If the misfit is larger than 15, the system issues a warning and prompts the user to verify the reconstruction visually.

4 Performance and experience

The algorithm has been tested on more than 5000 images taken in many countries and on many types of films. Some of the images were 10 years old. The images also included men, and children down to the age of ten.

The system has high specificity: In 0.25% of the cases the system locates a wrong structure, for instance it selects metacarpals 1-2-3 instead of 2-3-4.

The system has high sensitivity: In 0.25% of the cases it reports no found object although it is a valid image of the three middle metacarpals.

In conclusion the system performs correctly on 99.5% of the images.

The system is delivered with a protocol for recording the radiographs, and if this is followed the performance is better than 99.5%.

The 0.25% wrong reconstructions all occurred in cases where the system did request a visual check (misfit>15). This request comes up in on average 3% of the cases. This is a big advantage of the device. The DXA scanners often require the user to correct the reconstruction manually, which requires trained people. With the ASM, our system is better suited to routine examination of many patients. The fact that 3% of the cases requires attention is well received by the users. Often these images are peculiar and the operator appreciates this awareness to unusual cases and potential problems. This is a very conspicuous reflection of the power of the underlying specific shape model, which adds values to the system.

5 A More Active Shape Model, MASM

The translation operator demonstrates that there are more information available from the line search at the marks than is used in standard ASM. In analogy with AAM one can formulate a more active shape model that includes all four Euclidean operators, and not only the single translation operator.

To formulate this model, the coordinates of the marks are denoted x, and the pose of the shape by the 4-vector t. The aligned marks x' are related to the unaligned marks through the Euclidian transformation:

$$x = T_t(x')$$

The shape PCA parameterises the variations of the aligned shapes in terms of the shape scores b_s

$$x' = x'_{\text{mean}} + P_s\, b_s$$

Here the columns of the matrix P_s are the first few eigenvectors of the shape variation PCA. MASM also searches for the edge along a linepiece at each mark, but does not alter x directly, because in MASM the dynamic variables are (b_s, t) rather than x. When the edges have been located, the signed distances s from the model marks to the image edges are collected in a vector s, which is then used to generate the update of (b_s, t) in analogy with AAM:

$$\delta b_s = B_s\, s$$
$$\delta t = B_t\, s$$

This model needs more training than ASM, which only learns the shape modes P_s and the edge signatures. In addition MASM needs to learn B_s and B_t by applying known perturbations to the training examples. Thus MASM contains more knowledge and is therefore likely to act stronger. The good results with the simple translation operator used in the densitometry device shows that the use of this extra knowledge can be crucial in some applications.

Finally, AAM can be run also with (b_s, t) as dynamical variables (see section 9.2 of [6]), and updating these variables based on both edge errors (s) and intensity pattern errors (δg in the AAM notation) leads to an algorithm that matches both edges and intensities simultaneously.

6 Conclusion

ASM is useful for machine vision, but in practice a lot of development needs to be done to achieve a high performance in the recognition: 99.5%, rather than say 97%.

ASM is appropriate if a high degree of "understanding" is desirable. The extensive training of the models on a large biological material makes the model behave in an "experienced" manner. In the context of a medical device, this is an attribute of much value. The fact that the system can demonstrate its high knowledge in such a conspicuous way justifies a good measure confidence in the system, that it can demonstrate its high knowledge content in such a conspicuous way. It also allows the device to be operated by non-experts, which is essential in routine clinical work.

The approval by the USA FDA and the sale of more than 100 units means that this device is now an important contribution to the prevention and treatment of osteoporosis.

A theoretical spin-off from this work is the translation operator, which originated as an ad hoc solution to a limitation of ASM on the particular shapes being reconstructed. It has been generalised to the More Active Shape Model that helps bridge the ASM and AAM models and opens for more powerful models that fit both edge and intensity patterns in the recognition of deformable objects.

References

[1] T. F. Cootes, C. J. Taylor: Active Shape Models – 'Smart Snakes', in Proc British Machine Vision Conference, Springer-Verlag, 1992.

[2] T. F. Cootes, A. Hill, C. J. Taylor, J. Haslam: The Use of Active Shape Models for Locating Structures in Medical Images, Image and Vision Computing Vol. 12, No. 6, (1994)

[3] N.D. Efford: Knowledge-based segmentation and feature extraction of Hand Radiographs, Report 94.31 University of Leeds, October 1994.

[4] P. P. Smyth, C. J. Taylor J. E. Adams: Vertebral Shape: Automatic Measurements with Active Shape Models, Radiology, Vol. 211, No. 2, (1999).

[5] 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

[6] T. F. Cootes, C. J. Taylor: A detailed report about Active Shape Models and Active Appearance Models, on Tim Cootes' home page (Postscript), (2000).