BMVC99

Using Wavelets for Compression and Multiresolution Search with Active Appearance Models

C. B. H. Wolstenholme and C. J. Taylor

Division of Imaging Science and Biomedical Engineering,
University of Manchester,
Manchester M13 9PT, U.K.
cbw@sv1.smb.man.ac.uk
http://www.isbe.man.ac.uk

Abstract

Active Appearance Models (AAMs) provide a method of modelling the appearance of objects in images and locating them automatically. Although the AAM approach is computationally efficient, the models used to search unseen images for the structures of interest are large - typically the size of 100 images. This is perfectly practical for most 2-D images, but is currently impractical for 3-D images. We present a method for compressing the model information using a wavelet transform. The transform is applied to a set of training images in a shape-normalised frame, and coefficients of low variance across the training set are removed to reduce the information stored. An AAM is built from the training set using the wavelet coefficients rather than the raw intensities. We show that reliable image interpretation results can be obtained at a compression ratio of 20:1, which is sufficient to make 3-D AAMs a practical proposition.

1 Introduction

We have recently described Active Appearance Models (AAMs) which provide a generic approach to modelling the shapes and grey-level appearance of the structures of interest in a class of images and for locating them automatically by matching the models to unseen images [3]. So far we have only described 2-D applications of AAMs [7, 2]. In principle AAMs could be extended straightforwardly to 3-D. However, although the approach is computationally efficient, the models are large - typically the size of 100 images. This is perfectly practical for most 2-D images, but is currently impractical for 3-D images.

Wavelet transforms provide a useful approach to image compression [9, 1, 6]. By transforming an image, they enable relatively high compressions with very little degradation to the original image. In medical images all information is potentially important, and the near lossless nature of wavelet compression lends itself well to medical image compression [10].

BMVC 1999 doi:10.5244/C.13.47

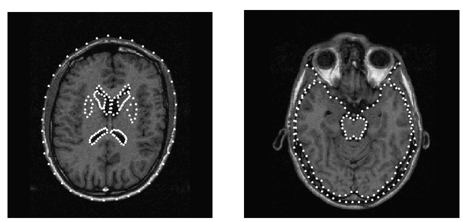
By combining the image search effectiveness of the AAM with the compression capabilities of wavelets, we hope to develop a 3D image search algorithm which is both efficient and robust.

2 Active Appearance Models

In this section we provide a brief outline of the standard AAM approach, and describe two applications used to evaluate our new approach. For a more comprehensive description of the AAM algorithm, see [3].

2.1 Modelling Image Appearance

Active Appearance Models are generated using a statistical analysis of the shape and texture variation over a training set of images. Rather than model the complete image, a region of interest is first labelled using a set of landmark points that describe the shape of the labelled objects in each image. Each example in the training set is labelled with the same number of points marking out the same structures. Figure 1 shows two examples of images labelled with landmark points.



(a) (b)

Figure 1: Two examples of an MR brain slice labelled with landmark points. One around the ventricles, the caudate nucleus, the lentiform nucleus, and the outside of the skull (a), and the other around the brain stem, brain hull and inside of the skull (b)

The variation in shape across the set is described by applying Principle Component Analysis (PCA) to the landmark points, resulting in a Point Distribution Model (PDM). Full details of this method can be found in [5]. Any valid example of the shape modelled can then be approximated using:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P_s}\mathbf{b_s} \tag{1}$$

where $\bar{\mathbf{x}}$ is the mean shape vector, $\mathbf{P_s}$ is a set of orthogonal modes of variation, and $\mathbf{b_s}$ is a vector of shape parameters.

To build a statistical model of the grey-level appearance we sample points within the convex hull of landmark points. These samples are then warped to the mean shape using a triangulation algorithm. Grey level information $\mathbf{g_{im}}$ is then sampled from this shape-normalised image. By applying PCA to the normalised data, we obtain a linear model that can approximate valid examples of grey-level appearance:

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P_g b_g} \tag{2}$$

where $\bar{\mathbf{g}}$ is the mean shape-normalised grey-level vector, $\mathbf{P_g}$ is a set of orthogonal modes of intensity variation, and $\mathbf{b_g}$ is a set of grey-level parameters.

By varying the vectors $\mathbf{b_s}$ and $\mathbf{b_g}$, the shape and grey-level of any example can be approximated. There may exist some correlation in the variances of shape and grey-level, so a further PCA can be applied to the data, concatenating the vectors, and obtaining a model of the form:

$$\begin{pmatrix} \mathbf{W_s b_s} \\ \mathbf{b_g} \end{pmatrix} = \mathbf{b} = \begin{pmatrix} \mathbf{Q_s} \\ \mathbf{Q_g} \end{pmatrix} \mathbf{c} = \mathbf{Qc} \tag{3}$$

where $\mathbf{W_s}$ is a diagonal matrix of weights for each shape parameter, correcting for the difference in units between the shape and grey-level models, $\mathbf{Q}$ is a set of orthogonal modes of appearance variation, and $\mathbf{c}$ is a vector of appearance parameters.

The linearity of the resulting model enables us to express shape and grey-levels as functions of $\mathbf{c}$

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P_s W_s Q_s c} \ , \ \mathbf{g} = \bar{\mathbf{g}} + \mathbf{P_g Q_g c} \tag{4}$$

A synthetic image can be generated for a given $\mathbf{c}$ by first generating the shape-free grey-level image, $\mathbf{g}$, and warping it to match the known control points described by $\mathbf{x}$.

2.2 Appearance Models of the Brain

Two sets of images were labelled with landmarks to make up the training sets for two appearance models. Both used slices from T1 weighted MR images of the brain. The training set for the ventricle model ((a) in Figure 1) contained 36 labelled images each with 163 landmark points, and the brain stem model ((b) in Figure 1) contained 42 images in its training set, each with 145 landmark points.

The effect of varying parameters in the vector $\mathbf{c}$ in each model is shown in Figure 2, where the first two parameters, c_1 and c_2, are varied by ± 2 standard deviations.

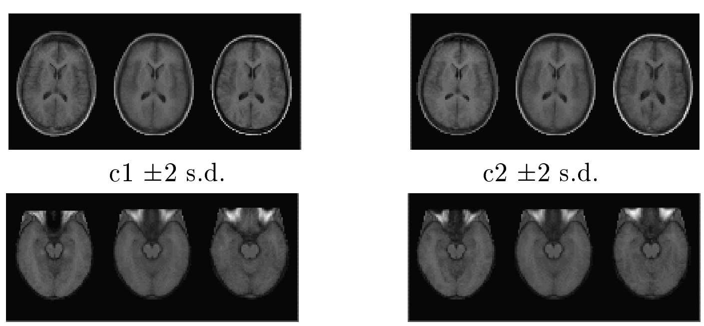


Figure 2: The first two modes of appearance variation in the two models

2.3 Image Search

To complete the tests for the standard model, an AAM using both models was built and tested in a model search.

The AAM search algorithm tries to minimise the difference in the model frame between a sample, $\mathbf{g_s}$, from an image being searched, and a synthesised image, $\mathbf{g_m}$, by varying the vector of appearance model parameters $\mathbf{c}$ which, for notational simplicity, is taken to include parameters for position, orientation and scale. This difference is given by:

$$\delta \mathbf{g} = \mathbf{g_s} - \mathbf{g_m} \tag{5}$$

Remember the $\mathbf{c}$ parameter also vary the shape, which affects the way the sampled image $\mathbf{g_s}$ is acquired. A linear relationship between $\delta \mathbf{g}$ and $\delta \mathbf{c}$, the change in $\mathbf{c}$ required to minimise $|\delta \mathbf{g}|$, is learnt during a second training phase:

$$\delta \mathbf{c} = \mathbf{A} \delta \mathbf{g} \tag{6}$$

The matrix $\mathbf{A}$ is obtained by applying linear regression using random displacements, $\delta \mathbf{c}$, from an image generated using the mean shape, pose and contrast. Equation (6) can then be used in an iterative matching algorithm by measuring the difference, δg, between an image generated by the current model parameters and predicting the change, $\delta \mathbf{c}$, required to minimise $deltag$. Full details of this can be found in [3].

Figure 3 shows results of AAM image search in each of the models described above. Quantitative results are given in Section 3.4.

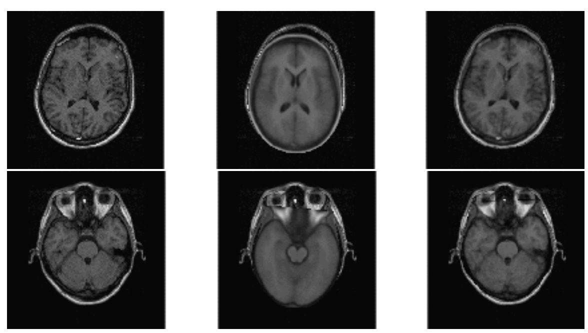


Figure 3: Search results for the two brain models, showing the original image, the start position of the model, and the search result

3 Compressing the Model

3.1 Wavelets

Wavelet analysis can be carried out using the Fast Wavelet Transform (FWT) algorithm developed by Stéphane Mallat [8]. This involves the decomposition of a signal by passing it through high-pass and low-pass filters. The wavelet coefficients are generated using a convolution of the high-pass and low-pass filters with the signal in turn. The low-pass filter gives rise to the *approximation*, which approximates the original signal, and the high pass filter gives rise to the *details*, which represent the minor corrections to the approximation for the accurate reconstruction of the signal. To reduce the data resulting from filtering the image twice, *downsampling* is used, removing coefficients at every second location.

In 2D signals, details are required in multiple directions - one horizontal, one vertical, and one diagonal. This is implemented using an extra bank of filters through the second dimension. Figure 4 shows this algorithm diagrammatically. To reconstruct the signal from the wavelet coefficients we need to reverse the pro-

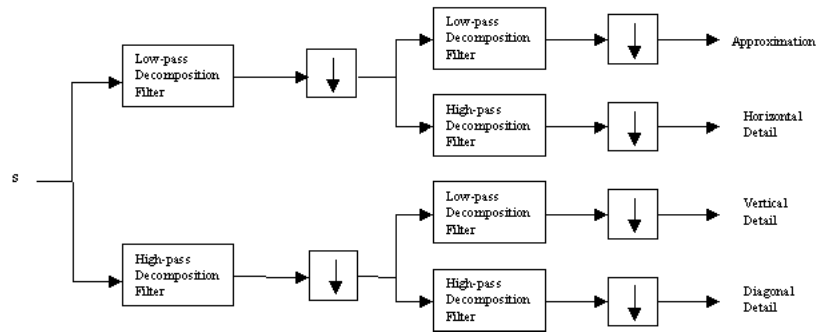


Figure 4: 2D wavelet decomposition. The rows of the signal, s, are passed through high- and low-pass filters, then downsampled. The resulting columns of the signal are then passed through further filters to give the approximation and three sets of details.

cess and replace the decomposition filters with reconstruction filters. Upsampling inserts zeros at every second position to correct for the downsampling during decomposition.

In the standard FWT, the filtering process can be applied to the approximation resulting from a previous pass, creating a multi-level tree. Further decomposition can be achieved using wavelet packet analysis. This is similar to the method described above, except that it not only decomposes the approximations at each level, but also the details , leading to a binary (one-dimesional decomposition) or quad- (two-dimensional decomposition) tree (Figure 5). As the details are the higher frequencies, which only add minor information to a signal or an image, then changing the values of small details has little effect on the reconstructed image. This is the basis of wavelet compression. In standard wavelet compression, a threshold is set, and detail coefficients falling below that threshold are zeroed. The zeroed values can then be removed to reduce the data, provided a scheme for their replacement is devised.

3.2 Compression Over an Image Set

When building an appearance model, a training set of images is used, each marked with the landmark points. After the shape model has been trained using the landmarks, the grey-level texture model is built. This is done by taking a convex

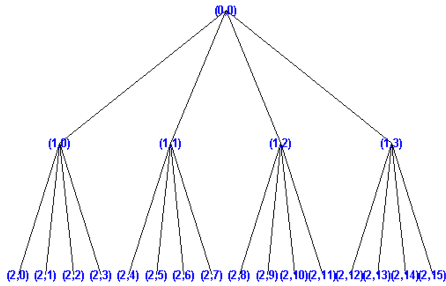


Figure 5: A wavelet tree to two levels. Each node holds a set of coefficients giving a detail or approximation of the previous level

hull around the landmarks, sampling from the training image, and warping the sampled data to the mean shape. This puts the grey-level samples in a shape free frame, as shown in Figure 6. In this shape free frame all the images are similar,

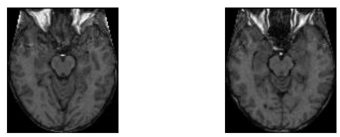


Figure 6: The shape free samples taken from the first and second images in the brain stem model's image set

so this is an ideal place to compress the images, since a standardised vector can be held for all the images detailing the decompression information. In the case of standard wavelet compression, zeroed coefficients that match throughout the image set, after the threshold is applied, can be zeroed. A vector containing the positions of the zero columns could be stored, and the zeros replaced at reconstruction time. The problem with this method is that the zeroed coefficients throughout the set rarely fall into line, giving only low rates of compression.

A better scheme, giving much higher rates of compression, is to remove the coefficients that show the least variance throughout the image set. Rather than zero these coefficients, they can be set to their mean to preserve any non-zero, but invariant values. A vector is then held which contains the means of all the positions that have been removed. This vector can be used during reconstruction to replace the mean values. The resulting vector can be passed through the wavelet reconstruction to recover the shape-free image, with minimised loss of quality.

This method differs from the standard wavelet compression method in that compression can be applied to the approximation as well as the detail without much degradation of the images. This is possible since the coefficients are set to their mean rather than being zeroed. The wavelet decomposition is still important to the method as most of the change is still in the detail coefficients. This is due to the fact that, in the 2D analysis, the details outweigh the approximations by 3:1 at the first level of the analysis, and this increases at the lower levels. We have used three levels of decomposition in our experiments.

Finally, an Appearance Model can be trained on the set of compressed wavelet coefficients instead of the original image information, reducing the size of the models.

3.3 Improving the AAM Search with Wavelets

A multiresolution version of the AAM algorithm outlined in Section 2 has been described previously and shown to improve both speed and robustness [4]. The method requires extra training of the Appearance Model at each resolution at which the search is to be carried out. As the aim of the wavelet tranform is to compress the Appearance Model, the standard multiresolution algorithm could not be applied. One property of the wavelet decomposition, however, is the downsampling and repeated decomposition of the approximations. These approximations lie in the left hand nodes of the wavelet tree (Figure 5), such that node **n,0** is the approximation at level **n**. The images at these nodes can be reconstructed, and represent a multi-level decomposition of the approximations as shown in Figure 7

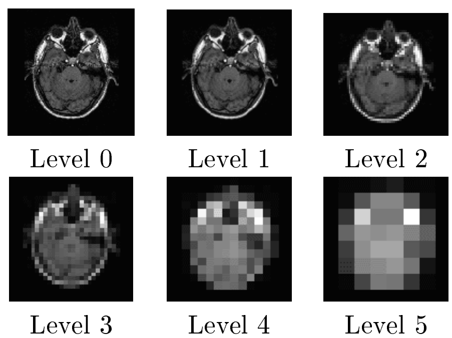
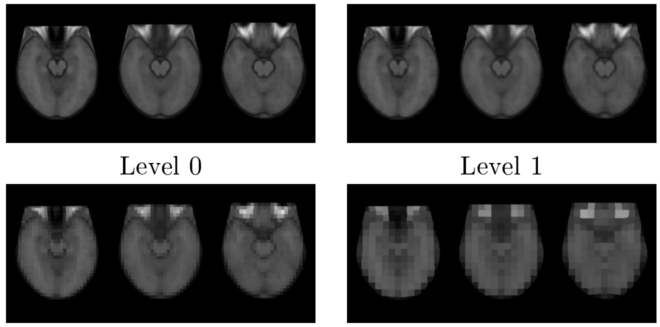


Figure 7: Images recovered from the different levels of approximations in the wavelet tree, showing their multiresolution property

As the model is decomposed to three levels, then we can reconstruct at each level to represent that resolution of the model. Figure 8 shows the Appearance Model representation at four resolutions (level 0 being that of the normal Appearance Model). The AAM can also be trained using the multi-level model. The mean model is reconstructed at each level, and random model displacements are reconstructed at the same level to train a regression matrix at that level. This is repeated at each level in the tree. In image search, a wavelet transform identical to that of the model is applied to the image to be searched, and its approximations are reconstructed at each level in turn. The model then searches at each level from the lowest resolution to the highest, with the result at each level used as the starting displacement for the next.

3.4 Compressed Appearance Models of the Brain

Compressed Appearance Models were built using the sets of training images labelled as shown in Figure 1, and three levels of Haar wavelet decomposition. The

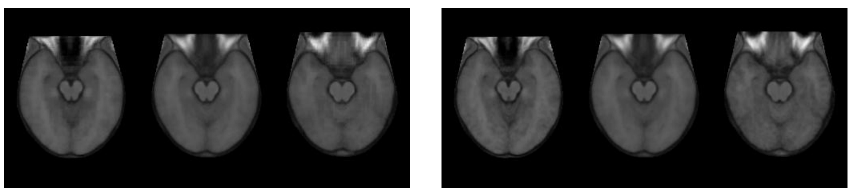


Level 0 Level 1

Level 2 Level 3

Figure 8: The appearance model reconstructed at the different levels of approximations

models were tested at 75%, 90% and 95% compression, giving compression ratios of 4:1, 10:1 and 20:1 respectively. The first mode of shape variation resulting from the brain stem model at 20:1 compression is shown in Figure 9 together with the uncompressed model.



(a) (b)

Figure 9: The first mode of variation of the model of the brain stem compressed to 20:1 (a) compared to the normal model (b)

Since the models built using wavelet coefficients could still represent the appearance of the images, and could be reconstructed with reasonable accuracy, AAMs of the models were built to test the compressed model's ability in image search. Only the models at 10:1 and 20:1 compression were used for this experiment. 'Leave-all-in' and 'leave-one-out' tests were used to give estimated upper and lower bounds on search accuracy. Figure 10 shows an example result from the tests using the 20:1 compressed model with the multiresolution search, and Table 1 shows the mean point-to-point errors between the search results and the original training landmarks for the compressed and non-compressed models. All the models tested gave similar results in the leave-all-in tests, showing that the compression had little effect on the AAMs search capabilities on previously seen images. In the leave-one-out tests, however, while the non-compressed and compressed models using the standard search again gave similar results, the multilevel search greatly improved the model's results. The non-compressed and 20:1 compressed models gave mean errors of 16.33 and 16.89 pixels respectively at their worst (and these failed to locate the brain), while the multilevel model at 20:1

compression resulted in a worst mean error of only 6.35 pixels, which located the brain. The best mean error for the multilevel model in the leave-one-out tests was also better at 1.72 pixels compared to the non-compressed (1.90 pixels) and 20:1 compressed (1.94 pixels) models.

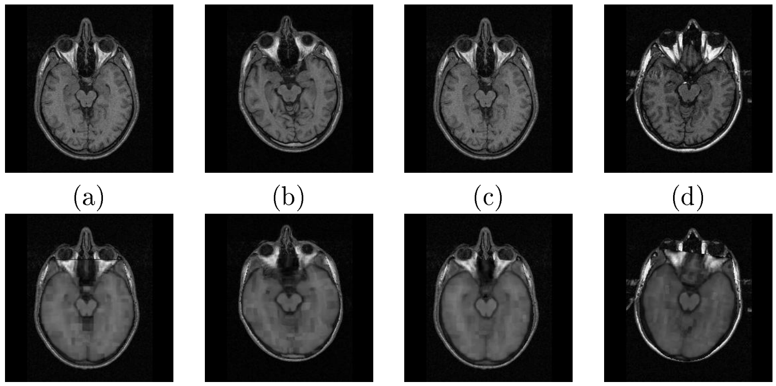


Figure 10: The search results of the 20:1 compressed brain stem model from 'leave-all-in' (best (a) and worst (b)) and the 'leave-one-out' (best (c) and worst (d)) tests using the multilevel search. The top row are the original images, and the bottom are the search results

Compression	Leave-all-in			Leave-one-out		
	% Success	*% Fail*	*P to P*	*% Success*	*% Fail*	*P to P*
None	100	0	3.03	92.86	7.14	4.47
20:1	100	0	3.11	90.48	9.52	4.78
Multi-res 20:1	100	0	3.25	100	0	3.89

Table 1: Search results from the leave-all-in and leave-one-out tests. The mean Point-to-Point error is measured from the reconstructed shape to the original landmarks in pixels.

4 Discussion and Conclusions

We have demonstrated a method for compressing the amount of data required to describe an Appearance Model, while still retaining the necessary information for using the model successfully in the AAM search algorithm. The method uses wavelets to place the shape-free images used by the appearance model into a frame where compression can be carried out with minimal degradation to the images. The compression is achieved by removing the coefficients that vary least throughout the training set within the wavelet space. We have shown that the model is still effective when used as an Active Appearance Model in image search.

The model loses detail as the compression increases, but it still searches accurately, and once the object is located in an image, the more detailed information can be retrieved directly from the image.

It is a small step from our current results to accurately locating brains in 2D, a process that may be used to bootstrap some existing 3D brain stripping algorithms. However, the reason for studying compression is to enable us to extend the AAMs to 3D. If AAM search can be used for 3D MR images of the brain, then this could prove an effective, fully automatic method for brain stripping and structure segmentation, as well as proving useful in solving other imaging location and segmentation problems.

References

[1] C K Chui. *Wavelet Analysis and Its Applications*. Academic Press, London, 1992.

[2] T F Cootes, C Beeston, G J Edwards, and C J Taylor. A unified framework for atlas matching using active appearance models. In *Press*, 1999.

[3] T F Cootes, G J Edwards, and C J Taylor. Active appearance models. In *Proceedings of the European Conference on Computer Vision*, volume 2, pages 484–498. Springer, 1998.

[4] T F Cootes, G J Edwards, and C J Taylor. A comparative evaluation of active appearance model algorithms. In *Proceedings of the British Machine Vision Conference*, volume 2, pages 680–689. BMVA Press, 1998.

[5] T F Cootes, C J Taylor, D H Cooper, and J Graham. Image search using trained flexible shape models. *Advances in Applied Statistics*, pages 111–139, 1994.

[6] I Daubechies. *Ten Lectures on Wavelets*. CBNS-NSF regional conference series in applied mathematics, Philadelphia, 1992.

[7] G J Edwards, C J Taylor, and T F Cootes. Learning to identify and track faces in image sequences. In *8^{th} British Machine Vision Conference*, pages 130–139. BMVA Press, 1997.

[8] S Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[9] G Strang and T Nguyen. *Wavelets and Filter Banks*. Wellesley - Cambridge Press, Wellesley, 1997.

[10] A Zandi, J Allen, E Schwartz, and M Boliek. Crew: Compression with reversible embedded wavelets. In *IEEE Data Compression Conference*, pages 212–221, 1995.