Improving Appearance Model Matching Using Local Image Structure

I.M. Scott, T.F. Cootes, C.J. Taylor

Imaging Science and Biomedical Engineering,
University of Manchester, Manchester, M13 9PT, UK.
ian.m.scott@stud.man.ac.uk

Abstract. We show how non-linear representations of local image structure can be used to improve the performance of model matching algorithms in medical image analysis tasks. Rather than represent the image structure using intensity values or gradients, we use measures that indicate the reliability of a set of local image feature detector outputs. These features are image edges, corners, and gradients. Feature detector outputs in flat, noisy regions tend to be ignored whereas those near strong structure are favoured. We demonstrate that combinations of these features give more accurate and reliable matching between models and new images than modelling image intensity alone. We also show that the approach is robust to non-linear changes in contrast, such as those found in multi-modal imaging.

1 Introduction

This paper builds on Cootes's *et al.*[2, 5] work on constructing statistical appearance models and matching them to new images using the Active Appearance Model (AAM) search algorithm. When building models of the appearance of objects it is advantageous to choose a representation of the image structure which can capture the features of interest in a way that allows a reliable comparison between model and image, and is invariant to the sorts of global transformation that may occur. For instance, when building statistical appearance models[2, 18] it is common to represent the image texture by a vector of intensity values sampled from the image, normalised by a linear transform so as to be invariant to global changes in brightness and contrast. By sampling across the whole region, all image structures are represented and all pixels treated equally (though the statistical analysis will then typically favour pixels in some regions over others, as dictated by the data.) However, models based on raw intensity tend to be sensitive to changes in conditions such as imaging parameters or biological variability. Thus models built on one data set may not perform well on images taken under different conditions. Also, intensity models do not explicitly distinguish between areas of noisy flat texture and real structure, and thus may not lead to accurate fitting in AAM search.

Edge based representations tend to be less sensitive to imaging conditions than raw intensity measures. Thus an obvious alternative to modelling the intensity values directly is to record the local image gradient in each direction

at each pixel. Although this yields more information at each pixel, and at first glance might seem to favour edge regions over flatter regions, it is only a linear transformation of the original intensity data. Where model building involves applying a linear Principal Component Analysis (PCA) to the samples, the resulting model will be almost identical to one built from raw intensities, apart from some effects around the border where computing the gradients includes some background information into the model.

In this paper we present a new representation. Rather than just recording the intensities at each pixel, we record a local structure tuple. It is useful to think about the rest of this work as using *texture preprocessors* which take an input image, and produce an image of tuples representing various aspects of local structure. This local structure tuple can include such things as edge orientation, corner strength, etc. When sampling the image to produce a texture vector for a model, instead of sampling n image intensity values from the original image, we sample all the values from each m-tuple at n sample locations, to produce a texture vector of length nm.

The local structure descriptors that we have used are gradient orientation (which was first discussed in a previous paper[4],) corner and edge strength. We demonstrate that using all of these measures in the texture preprocessor gives improved AAM matching accuracy and reliability when compared to intensity texture AAMs alone. We demonstrate that these improvements are statistically significant. We also show that the new approach can deal with images subject to strong non-linear changes in contrast, as found in multi-modal imaging.

2 Background

Eigen-faces[18] model the statistics of the intensities in a region of an image, and have been widely used for object location and recognition. Moghaddam and Pentland generalised this to include models of smoothed canny edge images[14]. In the image registration community, edge maps are widely used[17]. However, they tend to use either linearly normalised gradients or squared gradients, or non-maximally suppressed edges (all pixels other than those thought to be exactly on the edge are set to zero).

Edge orientation images have been used for face recognition by Hond and Spacek[7], who created histograms of edge orientation over regions and obtained good results. One of our structure descriptors can be thought of as a weighted version of edge orientation, in which strong edges are given more weight than weak edges. Rather than use a histogram we model the edge structure at every pixel.

Bosch *et al.*[1] have used a non-linear normalisation step on intensities in echo-cardiograms as pre-processing before building an appearance model. The intention is to modify the strongly non-gaussian noise statistics of such images into more normal shaped distributions. This gives significantly improved results. However, this was on intensities, not structure descriptors, and the approach described may not be optimal when applied to structure descriptors.

Several authors have attached feature detectors to points on an Point Distribution Model (PDM). This PDM can been automatically generated created using elastic variation of a single image[10]. A manually trained, but statistically learnt PDM can be used with profiles[3], and Gabor jets[11]. In all these approaches there is no dense model of texture, and the feature detector location, and effect on the shape model, has been set by humans rather than learnt.

Kittler *et al.*[9] demonstrated that different types of image normalisation could have a significant effect on a face verification task. They found that histogram normalisation tended to perform well over a range of experiments.

3 Statistical Models of Appearance

An appearance model can represent both the shape and texture variability seen in a training set. The training set consists of labelled images, where key landmark points are marked on each example object. An appearance model can be thought of as generalisations of eigen-patches or eigen-faces[14, 18] in which, rather than represent a rigid region, we model the shape of the region and allow it to deform.

Given a training set we can generate statistical models of shape and texture variation using the AAM method developed by Cootes *et al.*[2]. The shape of an object can be represented as a vector $\mathbf{s}$ of the positions of the landmarks and the texture (grey-levels or colour values) as a vector $\mathbf{t}$. This texture is sampled after the image has been warped to the mean shape. The texture preprocessing described in this paper also takes place after the texture has been warped to the mean shape. The appearance model has parameters, $\mathbf{c}$, controlling the shape and texture according to

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{Q}_s\mathbf{c}$$
$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{Q}_t\mathbf{c}$$

where $\bar{\mathbf{s}}$ is the mean shape, $\bar{\mathbf{t}}$ the mean texture and $\mathbf{Q}_s$,$\mathbf{Q}_t$ are matrices describing the modes of variation derived from the training set.

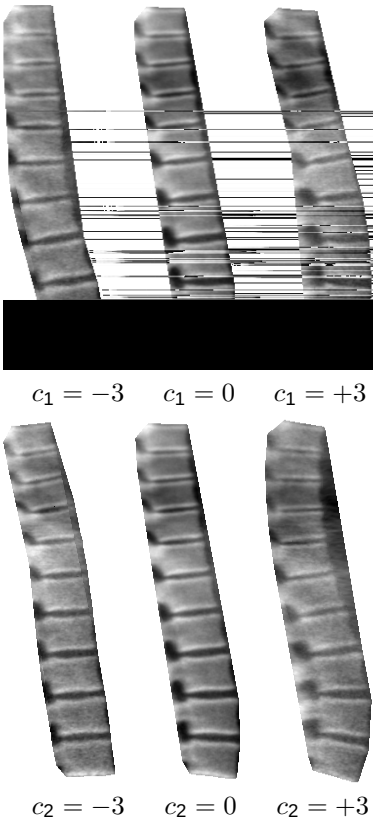

Fig. 1. Effect of varying first two parameters of a spinal X-ray appearance model, by ± 3 standard deviations from the mean.

An example image can be synthesised for a given $\mathbf{c}$ by generating a texture image from the vector $\mathbf{t}$ and warping it using the control points described by $\mathbf{s}$ (see figure 1.)

Such a model can be matched to a new image, given an initial approximation to the position, using the AAM algorithm[2]. This uses a fast linear update scheme to modify the model parameters so as to minimise the difference between a synthesised image and the target image. Appearance models and AAMs have been shown to be powerful tools for medical image interpretation [13, 1] and face image interpretation [2].

4 Local Structure Detectors

4.1 Non-linear transforms

As noted earlier, the texture preprocessor needs to be non-linear to make a significant difference to a linear PCA-based model. If we restrict the choice of preprocessor to those whose magnitude reflects the strength of response of a local feature detector, then it would be useful to transform this magnitude m into a reliability measure. We have chosen to use sigmoid function for this non-linear transform;

$$f(x) = \frac{m}{m + \overline{m}} \tag{1}$$

where $\overline{m}$ is the mean of the feature response magnitudes m over all samples. This function has the effect of limiting very large responses, preventing them from dominating the image. Any response significantly above the mean gives similar output. Also, any response significantly below the mean gives approximately zero output. This output behaves like the probability of there being a real local structure feature at that location.

4.2 Gradient Structure

The first local structure descriptor with which we have experimented is gradient orientation. Early work on non-linear gradient orientation is described in [4]. We calculate the image gradient $\mathbf{g} = (g_x \quad g_y)^T$ at each point giving a 2-tuple texture image for 2-d input images. The magnitude $|\mathbf{g}|$ can be transformed using equation 1, while preserving the direction. This is followed by the non-linear normalisation step to give

$$\mathbf{g_n} = \frac{(g_x \quad g_y)^T}{|\mathbf{g}| + \overline{|\mathbf{g}|}} \tag{2}$$

4.3 Corner and Edge Structure

We had observed that image corners were sometimes badly matched by gradient and intensity AAMs. Corners are well known as reliable features for corresponding multiple images[17], and in applications such as morphometry[15] accurate corner location is important in diagnosis.

Harris and Stephens [6] describe how to build a corner detector. They construct a local texture descriptor by calculating the Euclidean distance, or sum of square differences between an image patch and itself as one is scanned over the other. This local image difference energy E is zero at the patch origin, and rises faster for stronger textures.

$$E(x,y) = \sum_{u,v} [I(u+x, v+y) - I(u,v)]^2$$

To enforce locality and the consideration of only small shifts, they added a Gaussian window $w(u,v)$,and then made a first order approximation:

$$E(x,y) = \sum_{u,v} w(u,v) \left[x\frac{\partial I}{\partial u}(x,y) + y\frac{\partial I}{\partial v}(x,y) + O(x^2, y^2) \right]^2$$

Expanding the square-term gives

$$E(x,y) \quad = \quad Ax^2 + 2Cxy + By^2 \quad = \quad (x \;\; y)\mathbf{M}(x \;\; y)^T$$

where $w(u,v) = \exp -(u^2 - v^2)/2\sigma^2$, $\quad \mathbf{M} = \left(\begin{smallmatrix} A & C \\ C & B \end{smallmatrix}\right)$, $\quad A(x,y) = \left[\frac{\partial I}{\partial u}\right]^2 \otimes w$, etc.

The eigenvalues α,β of $\mathbf{M}$ characterise the rate of change of the sum of squared differences function as its moves from the origin. Since α and β are the principle rates of change, they are invariant to rotation. Without loss of generality, the eigenvalues can be rearranged so that $\alpha >= \beta$. The local texture at each point in the image can be described by these two values. As shown in figure 2, low values of α and β imply a flat image region. A high value of α and low value of β imply an image region flat in one direction, but changing in another, i.e. an edge. High values of both α and β imply a region that isn't flat in any direction, i.e. a corner.

At this point Harris and Stephens identified individual points of interest by looking for local maxima in $\det \mathbf{M} - k[\mathrm{tr}\,\mathbf{M}]^2$. We leave their approach here, except to note that useful measures derived from α and β can be found without actually performing an eigenvector decomposition, e.g. $\det(\mathbf{M}) = AB - C^2$

4.4 Developing measures of cornerness and edgeness

It would be useful to have independent descriptors of edgeness and cornerness. To force α and β into an independent form, we take the vector $(\alpha \;\; \beta)^T$ and double the angle from the α axis, as in figure 3. It is possible to calculate the cornerness, r, and edgeness, e, defined this way, without explicitly having to calculate an eigenvector decomposition. Note that our edgeness measure is different from the gradient measure, by being independent of edge direction.

$$\tan\theta = \frac{\beta}{\alpha} \quad \Rightarrow \quad \sin\theta = \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} \quad \text{and} \quad \cos\theta = \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}}$$

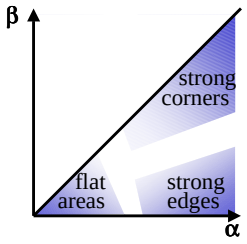

Fig. 2. How α and β relate to cornerness and edgeness.

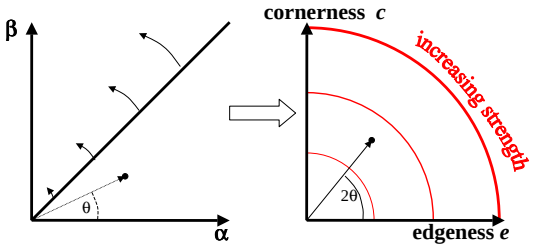

Fig. 3. Making cornerness independent of edgeness by doubling angle from axis.

$$\begin{aligned} r &= (\alpha^2 + \beta^2)\sin 2\theta \\ &= 2\det \mathbf{M} \\ &= 2AB - 2C^2 \end{aligned} \quad (3)$$

$$\begin{aligned} e &= (\alpha^2 + \beta^2)\cos 2\theta \\ &= \operatorname{tr}\mathbf{M}\sqrt{\operatorname{tr}^2\mathbf{M} - 4\det\mathbf{M}} \\ &= (A+B)\sqrt{(A-B)^2 + 4C^2} \end{aligned} \quad (4)$$

5 Experiment with Spinal X-Rays

We took a previously described[16] data set of low-dose Dual X-ray Absorptiometry (DXA) lateral scans of the spines of 47 normal women. The vertebrae from T7 to L4 were marked up under the supervision of an experienced radiologist — figure 4 shows an example. The images are 8-bit greyscale and roughly 140×400 pixels in size. Each vertebra is about 20-25 pixels tall.

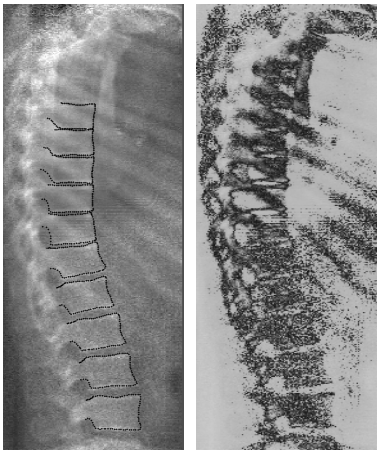

Fig. 4. A spinal DXA image with markup, and after multi-modal simulation.

Since we did not have a large data set, we performed leave-1-out experiments, by repeatedly training an AAM on 46 of the images and testing it on the remaining image. For each test image we performed 9 AAM searches starting with the mean shape learned during training, displaced by all combinations of $[-10, 0, +10]$ pixels in x and y. After the AAM search had converged we measured the distance from each control point on the AAM to the nearest point on the curve through the equivalent marked-up points. We calculated the mean of these absolute errors for each AAM search. Because of the even spacing of control points around each vertebra, this error will be approximately proportional to the total pixel overlap error.

This whole experiment was run for each of the following texture preprocessors:

Intensity Original AAM.

Sigmoidal gradient 2-tuple output $\mathbf{g_n}$ of sigmoidally normalised directed gradient (equation 2.)
Sigmoidal edge Sigmoidally normalised version of undirected edgeness e (equation 4.)
Sigmoidal corner Sigmoidally normalised version of cornerness r (equation 3.)
Sigmoidal corner and edge 2-tuple of the sigmoidally normalised cornerness and edgeness (r, e). (equations 3 and 4.)
Sigmoidal corner and gradient ...
Sigmoidal edge and gradient ...
Sigmoidal corner, edge, and gradient ...

In another experiment to simulate performance on multi-modal images, roughly half of the set of images were transformed by applying a bitonic pixel-value transfer function — see figure 4 for an example. The two groups were then merged, to give a set of 47 images. A leave-1-out experiment, similar to the previously described one, was then performed on this simulated multi-modal data set.

5.1 Results

The distribution of mean absolute errors for the $47 \times 9 = 423$ searches of the normal data set for three of the preprocessors is shown in figure 5. Each search was considered a success if the mean absolute point to curve error was less than 2 pixels. (The estimated repeatability of expert annotation is 1 to 1.5 pixels on this data.) Figure 6 summarises the results for all of the preprocessors. The results from the simulated multi-modal data set for the original "Intensity" and the "Sigmoidal corner, edge and gradient" AAMs are summarised in figure 7.

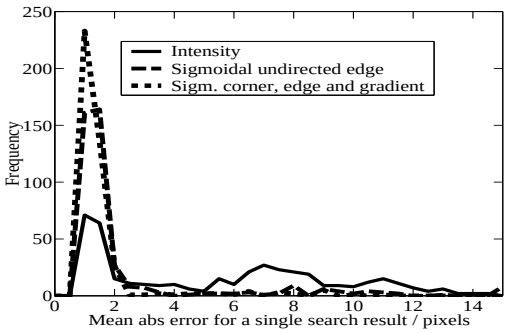

Fig. 5. Error spread between spinal AAM control points and the marked-up curves.

5.2 Statistics

It is not possible enough to show that the improvement is significant by simply comparing the means and standard deviations in figure 6, because the data is not normally distributed. Instead we use the percentage of successful results. If we classify the results as successes or failures according to the above test (section 5.1,) and count the number of successes, we should expect the result to be a binomially distributed random variable. When comparing two experiments, we need to show that any improvement in the percentage of successful results is statistically significant. To do this we must assume that there is an underlying distribution based on a probability of a single success of θ. After performing one experiment with n trials we get ny successes, and so estimate a probability

Fig. 6. Comparing point-to-curve errors (in pixels) for different spinal AAM texture preprocessors.

Texture Preprocessor	Searches	Point-Curve error			
	<2 pixels	mean	std	median	90%-ile
Intensity	35%	5.4	3.8	5.6	11.0
Sigmoidal gradient	40%	5.1	4.0	4.7	10.8
Sigmoidal edge	82%	2.4	3.1	1.4	6.5
Sigmoidal corner	75%	2.6	2.7	1.5	7.5
Sigmoidal corner and edge	81%	2.2	2.6	1.3	4.8
Sigmoidal corner and gradient	80%	2.1	2.2	1.2	1.9
Sigmoidal edge and gradient	85%	1.9	2.1	1.2	4.6
Sigmoidal corner, edge, and gradient	92%	1.5	1.4	1.2	1.8

Fig. 7. Comparing point-to-curve errors (in pixels) for simulated multi-modal spinal images

Texture Preprocessor	Searches	Point-Curve error			
	<2 pixels	mean	std	median	90%-ile
Intensity	7%	9.5	6.1	8.9	16.0
Sigmoidal corner, edge, and gradient	60%	3.4	3.8	1.6	9.3

of success y. We perform another experiment of m trials and get a probability of success x. We are interested in the probability of x being from the same distribution as y, having already measured y.

$$p(x|y) = \frac{p(x \cap y)}{p(y)}$$

Each of these probabilities depends on the parameter of the underlying binomial distribution $p(x|\theta)$, so we must marginalise θ away.

$$p(x|y) = \frac{\int_0^1 p(x \cap y|\theta)d\theta}{\int_0^1 p(y|\theta)d\theta} = \frac{\int_0^1 p(x|\theta)p(y|\theta)d\theta}{\int_0^1 p(y|\theta)d\theta}$$

where the binomial distribution is

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

It doesn't appear to be possible to find an analytic solution to these integrals, however we can use numeric integration. Figure 8 gives the p-values for each result, given a null hypothesis that a poorer performing experiment could have produced that result. It should be noted that because the 9 search tests per image can not be considered independent of each other, we based the significance calculation on a value $n = 47$.

Fig. 8. Probabilities (p-values) that an experiment could be a random result of a worse performing spinal experiment.

Texture Preprocessor	Result	$-\log_{10} p$-value given base result						
		35%	40%	75%	80%	81%	82%	85%
Intensity	35%							
Sigmoidal gradient	40%	0.5						
Sigmoidal corner	75%	4.7	3.9					
Sigmoidal corner and gradient	80%	5.6	4.8	0.6				
Sigmoidal corner and edge †	81%	6.1	5.3	0.8	0.5			
Sigmoidal edge †	82%	6.1	5.3	0.8	0.5	0.4		
Sigmoidal edge, and gradient	85%	6.7	5.8	0.9	0.6	0.5	0.5	
Sigmoidal corner, edge, and gradient	92%	9.5	8.5	2.2	1.7	1.4	1.4	1.2

† Note that the fraction of successful results is rounded down to the next lowest multiple of $1/n$ for p-value calculation, causing two rows with slightly dissimilar success rates to have identical p-values.

We can see that the large improvements between the "intensity" AAM and the various texture preprocessor AAMs are certainly significant. With the exception of the "sigmoidal gradient" preprocessor, the differences between the various texture preprocessors are not significant at the $\alpha = 0.01$ level. In the simulated multi-modal experiment, the improvement of the "Sigmoidal corner, edge, and gradient" preprocessor over the "intensity" AAM, is significant with $p = 5{\times}10^{-7}$.

6 Experiment with Faces

To add to the confidence of our results, we repeated the experiments with a face data set. We have a much larger database of face images. This enables us to build a face AAM using 100 images, and then test it on the independently collected extM2VTS[12] database. We have marked up 1817 images of 295 distinct subjects from this database. The remaining images in the database suffer from extreme motion blur, and interlace artifacts.

The raw "intensity", and "sigmoidal gradient" are repetitions respectively, of the NI and ES cases described by Cootes's and Taylor's paper[4]. That paper only used a 100 image subset of extM2VTS for testing, and did no statistical analysis of the results.

The results are tabulated in figure 9. The median and 90 percentile errors are given for all experiments. A search was considered successful when the mean absolute point to curve error fell below 5% of the inter-occular distance or 5 pixels in the Surrey dataset.

As described before we cannot use all the measurements because they are not independent. This is especially unfortunate with the faces experiments because we have a lot of data for which the tests should be largely if not completely

Fig. 9. Comparing point-to-curve errors for different facial AAM texture preprocessors.

Texture Preprocessor	Searches	Point-Curve error			
	<2 pixels	mean	std	median	90%-ile
Intensity	55.8%	5.4	2.9	4.6	9.0
Sigmoidal gradient	72.5%	4.5	2.0	3.9	7.2
Sigmoidal edge	68.8%	4.8	2.5	4.0	8.1
Sigmoidal corner	68.0%	4.8	2.3	4.0	7.8
Sigmoidal corner and edge	73.9%	4.5	2.3	3.8	7.6
Sigmoidal corner and gradient	83.6%	3.9	1.4	3.5	5.7
Sigmoidal edge and gradient	80.3%	4.1	1.7	3.6	6.2
Sigmoidal corner, edge, and gradient	83.7%	3.9	1.7	3.8	5.8

independent. In particular AAM searches of different images of the same person should be mostly independent. However, we again conservatively chose $n = 295$ as the number of strictly independent measurements.

Fig. 10. Probabilities (p-values) that an experiment could be a random result of a worse performing facial experiment.

Texture Preprocessor	Result	$-\log_{10} p$-value given base % rate						
		55.8	68.0	68.8	72.5	73.9	80.3	83.6
Intensity	55.8%							
Sigmoidal corner	68.0%	3.0						
Sigmoidal edge	68.8%	3.2	0.4					
Sigmoidal gradient	72.5%	5.0	1.0	0.8				
Sigmoidal corner and edge	73.9%	5.7	1.2	1.1	0.5			
Sigmoidal edge and gradient	80.3%	10.3	3.6	3.3	2.0	1.6		
Sigmoidal corner and gradient†	83.6%	13.5	5.5	5.1	3.4	2.9	0.9	
Sigmoidal corner, edge, and gradient†	83.7%	13.5	5.5	5.1	3.4	2.9	0.9	0.3

† See note in figure 8.

7 Discussion and Conclusion

We have shown that using descriptions of local structure for the texture model of an AAM significant improves the accuracy of the AAM search. Furthermore, the righthand mode of the distribution of "intensity" AAM results (figure 5) can be interpreted as convergence failures or false minima. The significant reduction in these failures using the various local structure preprocessors shows that we have also improved the reliability of AAMs.

The local structure descriptors are less dependent on the global or sub-global contrast effects caused by differing imaging parameters. The simulated multi-

modal spinal image experiment shows that the "intensity" AAM needs to devote so much variance to it's texture model to cope, that it fails to learn any useful information about the images. Comparing the results for the "Sigmoidal corner, edge and gradient" preprocessor in figures 6 and 7 shows that the severe image corruption has a relatively small effect on a local structure AAM.

Using all the sigmoidally-normalised local structure descriptors gives the best results on both the spine and face data. This suggests that it may be advantageous to add more local structure descriptors, including parameterised families of descriptors (provided that they all have magnitude based outputs, and distributions that are compatible with the non-linear sigmoidal function.) Potentially interesting families include the differential Gaussian invariants (which have been shown[19] to have high saliency in shape modelling applications) and complex wavelets[8]. Adding ever more local structure descriptors will inevitably lead to a case of diminishing returns. More work will be needed to determine if adding too many descriptors would significantly decrease performance, perhaps though increased numerical errors in the linear PCA learning.

It may be possible to normalise the statistics of any structure description (perhaps using the Cumulative Distribution Function (CDF) of the magnitudes over all or part of the image) and correctly combine non-magnitude (e.g. pixel intensity) descriptors with the sigmoidally-normalised descriptors. Even without fully solving this problem it should be straightforward to concatenate the sigmoidally-normalised descriptors from the different planes of co-registered multi-modal medical images.

This work should extend straightforwardly to 3D images. The calculation of gradient preprocessor extends trivially to 3D. The joint calculation of the corner, edge pair in 2D would extend to an analogous calculation of a plane, edge, corner triplet in 3D. However, it is difficult to do experiments with 3D volume AAMs due to the sheer size of such models.

Picking one local structure descriptor that is responsible for the majority of the improvement is possible in either data set. However, the "sigmoidal gradient" descriptor which is responsible for most of the performance improvement in the face data set, gives performance which is not significantly different from the "intensity" AAM in the spine data set. By providing the AAM training algorithm with all of the local structure descriptors, it can learn which descriptors are most useful, and adjust the importance of each descriptor on a sample by sample basis to get optimum performance.

Acknowledgements

We would like to thank Prof. Judith Adams and Martin Roberts for the data and useful discussions.

References

[1] H.G. Bosch, S.C.Mitchell, P.F.Boudewijn, P.F.Leieveldt, F.Nijland, O. Kamp, and M. Sonka. Active Appearance-Motion Models for endocardial contour detection

in time sequences of echocardiograms. In *SPIE Medical Imaging*, 2001.

[2] T.F. Cootes, G.J. Edwards, and C.J.Taylor. Active Appearance Models. *IEEE Transactions on Pattern Matching and Machine Intelligence*, 23(6):681–885, 2001.

[3] T.F. Cootes and C.J. Taylor. Active Shape Models - 'Smart Snakes'. In *British Machine Vision Conference*, pages 266–275. Springer-Verlag, 1992.

[4] T.F. Cootes and C.J. Taylor. On Representing Edge Structure for Model Matching. In *CVPR*, volume 1, pages 1114–1119, 2001.

[5] G. Edwards, Chris J. Taylor, and Tim F. Cootes. Interpreting face images using active appearance models. In *International Conference on Automatic Face and Gesture Recognition*, pages 300–5, 1998.

[6] C. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[7] D. Hond and L. Spacek. Distinctive Descriptions for Face Processing. In *British Machine Vision Conference*, Colchester, 1997.

[8] N. Kingsbury. Image processing with complex wavelets. *Philosophical Transactions of the Royal Society, A*, 357(1760):2543–60, 1999.

[9] J. Kittler, Y.P. Li, and J. Matas. On matching scores for LDA-based face verification. In *British Machine Vision Conference*, volume 1, pages 42–51. BMVA Press, 2000.

[10] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, and W . Konen. Distortion Invariant Object Recognition In The Dynamic Link Architecture. *IEEE Transactions On Computers*, 42(3):300–311, 1993.

[11] Stephen J. McKenna, Shaogang Gong, Rolf P. Wrtz, Jonathan Tanner, and Daniel Banin. Tracking Facial Feature Points with Gabor Wavelets and Shape Models. In *International Conference on Audio-Video Based Biometric Person Authentication*, pages 35–43, 1997.

[12] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSdb: The Extended M2VTS Database. In *2nd Conf. on Audio and Video-based Biometric Personal Verification*. Springer Verlag, 1999.

[13] S.C. Mitchell, P.F Boudewijn, P.F.Lelievedt, R.J. van der Geest, H.G. Bosch, J.H. Reiber, and M. Sonka. Time continuous segmentation of cardiac MR image sequences using Active Appearance Motion Models. In *SPIE Medical Imaging*, 2001.

[14] Baback Moghaddam and Alex Pentland. Probabilistic Visual Learning for Object Representation. *IEEE Transactions on Pattern Matching and Machine Intelligence*, 19(7):696–710, 1997.

[15] National Osteoporosis Foundation Working Group on Vertebral Fractures. Assessing vertebral fractures. *Journal of Bone and Mineral Research*, 10(4):518–23, 1995.

[16] Paul P. Smyth, Christopher J. Taylor, and Judith E. Adams. Vertebral Shape: Automatic Measurement with Active Shape Models. *Radiology*, 211:571–578, 1999.

[17] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall, 1998.

[18] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[19] K. Walker, T.F. Cootes, and C.J. Taylor. Locating Salient Object Features. In *British Machine Vision Conference*, pages 463–472, 1998.