Double Contour Active Shape Models

Matthias Seise[1], Stephen J. McKenna[1], Ian W. Ricketts[1]
and Carlos A. Wigderowitz[2]
[1]Division of Applied Computing, University of Dundee, UK
[2]Orthopaedic & Trauma Surgery, Ninewells Hospital, Dundee, UK
{mseise,stephen,ricketts}@computing.dundee.ac.uk, cawigderowitz@lineone.net
www.computing.dundee.ac.uk/projects/Vision

Abstract

Statistical shape models are often learned from examples based on landmark correspondences between annotated examples. A method is proposed for learning such models from contours with inconsistent bifurcations and loops. It is evaluated on the task of segmenting tibial contours in knee radiographs. Results are presented using various features, distance weighted $K-$nearest neighbours and differing eigenspace shape constraints.

1 Introduction

Statistical models of shape based on point distribution have enjoyed considerable success, particularly for segmentation, tracking and recognition of biological shape variation, e.g. face and medical image analysis. The original active shape model formulation and most recent methods based upon it rely on explicit inter-image correspondence being established between landmark points. These points often lie on identifiable contours in the images, their positions being determined either manually or (semi-)automatically [6].

Consider the image contours annotated in Figure 1. Shown are four examples from a lipreading application and four from a radiographic image analysis application. In both cases, contours can contain loops. Furthermore, the number of loops and the positions of the bifurcation points relative to the object's image projection vary in a complex way. Corresponding landmarks cannot be straightforwardly identified in these images. The use of bifurcation points as landmarks, for example, leads to undefined correspondence matches and unmeaningful variation. An alternative approach would be to treat each contour as

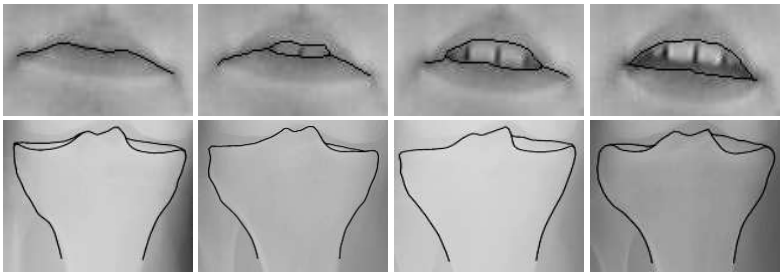


Figure 1: Shape examples annotated with looping contours. Top: a lipreading application. Bottom: radiographs of the tibia. Note that the number of loops and their positions vary.

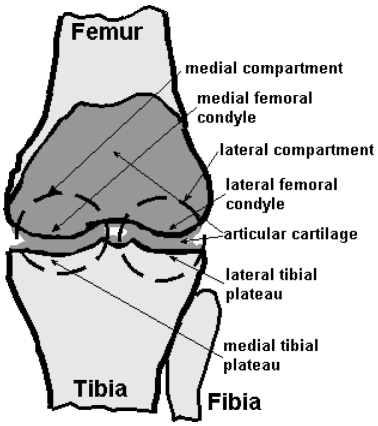


Figure 2: Anatomy of the knee.

Figure 3: Clinical x-ray (enhanced).

two contours, each sharing the two endpoints but taking inner and outer paths around the loops. Modelling these contours independently results in a poorly constrained search in which both contours often find the same side of a loop or only one of the contours localises a section where there is no loop. Even if the contour landmarks are concatenated to form a single landmark vector representation, the result will be two contours that in general will not be collinear where the expert annotation (ground-truth) would only indicate a single contour, a rather unsatisfactory state of affairs. In general, inner and outer contours would have differing landmarks on sections where the contours are collinear.

In this paper, the *double contour active shape model* (DCASM) is proposed. Double contours are parameterised in terms of one contour's landmarks along with a suitably constrained warp to the other contour. Correspondence is thus established between the contours. The method simplifies to a standard active shape model (ASM) in the case of a contour without bifuractions and can thus be considered to be a generalised ASM.

The paper also discusses other aspects of these models including the local appearance features used and the often overlooked effect of dimensionality on tests for shape plausibility. A distance weighted K-nearest neighbour (K-NN) method is proposed for local appearance matching and this is compared with the commonly used Mahalanobis distance.

The DCASM is evaluated here for the task of segmenting tibial contours in x-ray images of the knee, a useful step towards automated radiographic assessment of osteoarthritis (OA). Some background on this application is provided in the next Section.

2 Osteoarthritis

OA is the most common joint disease and the most common cause of disability in older people [8], resulting in significant economic costs for society. It is characterised by an imbalance of the synthesis and degeneration of the articular cartilage. In most cases of knee OA, the cartilage covering the tibial plateaux and the femoral condyles is being destroyed (see Figure 2). Two-dimensional x-ray imaging is the most widely used modality

for OA diagnosis and progression assessment. Cartilage is not visible in x-ray images so the primary radiographic sign used is the joint space between the lateral (medial) femoral condyle and the lateral (medial) tibial plateau. Joint space decreases as cartilage is destroyed. Other signs include the formation of osteophytes (bony spurs), cysts and subchondral sclerosis. Unfortunately, there are no standardised methods for measuring these signs. Physicians normally use only the minimum joint space width (JSW) to measure joint space narrowing, resulting in large inter- and intra-observer variation [2]. The use of active shape models to segment the femur and tibia in the knee joint in order to measure OA signs has not been attempted previously to the authors' knowledge.

3 Active Shape Models

Before introducing the DCASM, the standard ASM is first briefly described. Given a training set of S images in which the objects of interest are suitably annotated, statistical shape and appearance models can be estimated [5]. Correspondence must be established between training examples and this is often done manually by annotating landmark points. Alternatively, contours can be annotated and landmarks on the contours determined automatically based, for example, on minimum description length criteria [6]. The training examples are then aligned, typically using Procrustes analysis to determine translation, rotation and scale parameters that minimise distances between the corresponding landmarks in a least-squares sense. A shape is described by its N landmark points $\{(x_n, y_n)\}_{n=1}^{N}$. Each training example can be written as a $2N$ element vector $\mathbf{x}_s = (x_1^{(s)}, y_1^{(s)}, \ldots, x_N^{(s)}, y_N^{(s)})^\top$. Sample mean and covariance matrices are:

$$\overline{\mathbf{x}} = \frac{1}{S}\sum_{s=1}^{S}\mathbf{x}_s \qquad\qquad \mathbf{C} = \frac{1}{S-1}\sum_{s=1}^{S}(\mathbf{x}_s - \overline{\mathbf{x}})(\mathbf{x}_s - \overline{\mathbf{x}})^\top \tag{1}$$

Let $\Phi = (\phi_1|\phi_2|\cdots|\phi_D)$ denote the matrix whose columns are the D eigenvectors corresponding to the D largest eigenvalues $\lambda_1, \ldots, \lambda_D$ of $\mathbf{C}$. Any example of the training set, $\mathbf{x}_s$, can be approximated by

$$\mathbf{x}_s \approx \overline{\mathbf{x}} + \Phi\mathbf{b}_s \ , \tag{2}$$

where $\mathbf{b}_s$ is the D dimensional model parameter vector, computed by

$$\mathbf{b}_s = \Phi^\top(\mathbf{x}_s - \overline{\mathbf{x}}) \ . \tag{3}$$

The number D of eigenvectors to retain is usually calculated as the smallest D that satisfies $f_v \sum_{n=1}^{2N} \lambda_n \leq \sum_{d=1}^{D} \lambda_d$, where the parameter f_v is the proportion of the total variance of the data which can be explained, usually ranging between 0.900 and 0.995.

The appearance model describes the image structure around each landmark. It is usual to use fixed-length, one-dimensional profiles orthogonal to the contour. For each example and each landmark a fixed number of pixels on and to either side of the contour are sampled. Cootes and Taylor [5] originally proposed the use of normalised first order derivative profiles. Typically, the profile distribution is modelled as a multivariate Gaussian. Thus, the Mahalanobis distance can be used as a measure of the quality of fit of a profile.

Active shape model search is iterative and local. It is usually initialised with the mean shape and translation, rotation and scale parameters reasonably close to their 'true' values. At each iteration, points on and to either side of the contour along the normal direction are

considered. Profiles centred at each of these points are sampled and their Mahalanobis distances calculated. The landmark position is updated as the point with minimal Mahalanobis distance. After processing all landmarks, the closest plausible shape is found by projecting onto the eigenspace (Equation (3)). Plausible shapes are usually defined as those for which every shape parameter b_d is between $-3\sqrt{\lambda_d}$ and $3\sqrt{\lambda_d}$. Search is iterated a fixed number of times or until the shape model has converged. Search results can be improved if a multi-resolution, coarse-to-fine search is adopted with case appearance models learned for each resolution. The segmentation result at each resolution is used to initialise the search at the next resolution.

Several improvements to the standard ASM have been proposed. For example, more complex features characterising texture have been used for appearance modelling [12]. Active appearance models which model 2D appearance as well as shape variation using PCA are useful in applications such as human face analysis [4] although they often result in lower accuracy localisation of contours than ASM [3]. When a linear model of shape variation is inadequate, non-linear models have been used, e.g. [9, 10].

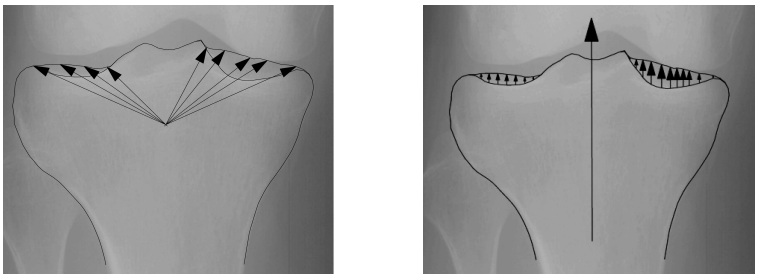
4 Double Contour ASM

In order to model double contours such as in Figure 1, each example is treated as two contours that share their endpoints and take inner and outer paths at bifurcations. One of these contours is used as a *reference contour* and the double contour is represented in terms of landmarks on this reference contour along with a constrained warp to the other contour. The warp defines corresponding landmarks on the other contour. In order for these to be positioned so as to form a good representation of shape, the warp must be suitably constrained. A warp suitable for certain radial shapes is to displace along the line between the centre of gravity of the reference contour $(\bar{x},\bar{y}) = \left(\frac{1}{N}\sum_{n=1}^{N} x_n, \frac{1}{N}\sum_{n=1}^{N} y_n\right)$ and the corresponding landmark (x_n, y_n) (see Figure 4(a)). Another possibility is displacement along the normals. Note, however, that both can yield inappropriate landmarking. In the case of the example applications in Figure 1, a suitable warp can be achieved by parallel displacement of the landmarks, i.e. by translating each of the N reference landmarks, (x_n, y_n), a distance δ_n in a shared direction θ. For the tibia, θ can be defined as an approximation to the dominant bone axis (see Figure 4(b)). This is because contour loops occur only at the tibial plateaux. Separations in this direction are always well defined. Specifically, the orientation of the line connecting the two bifurcation points for each loop in each training example is computed. The direction θ can then be set normal to the angular mean of these orientations. This leads to a suitable warp for the lipreading application, for example. In the case of the tibia where the loops can clearly be distinguished as medial or lateral, θ was set to the orientation of the angular bisector of the median lateral and median medial bifurcation lines.

Each example $\mathscr{S}_s$ is represented by a $3N$ element shape vector $\mathbf{x}_s$ and the dominant axis θ_s,

$$\mathscr{S}_s = (\mathbf{x}_s, \theta_s) \text{ where } \mathbf{x}_s = (x_1^{(s)}, y_1^{(s)}, \delta_1^{(s)}, \ldots, x_N^{(s)}, y_N^{(s)}, \delta_N^{(s)})^\top \tag{4}$$

The correspondence between the reference contour and the other contour is explicitly given by the direction of the dominant axis and the separation, δ_n, for each landmark point. Therefore, only the landmarks on the reference contours of each shape now need



(a) (b)

Figure 4: Two landmark warps: (a) radial and (b) parallel.

to be brought into correspondence. This can be achieved using the minimum description length method [6].

A naive approach to aligning two shapes would be to treat a vector $\mathbf{x}$ as defining a shape in $3D$ space and to align them in this $3D$ space. This does not work because the $3D$ transformations (rotation and translation) do not treat the separation dimension appropriately. Both contours should be taken into account during alignment. Furthermore, care must be taken to weight the effect of the two contours and the effect of landmarks on either side of a bifurcation equally. This is achieved by converting the double contour representation into a vector $\tilde{\mathbf{x}}$ of $2D$ image vectors for both the reference contour landmarks and those outer contour landmarks at which the separation in at least one example shape is non-zero. More formally, let $I = \{i_1, \ldots, i_M\}$ denote the set of indices where the separations are not always zero, thus $i_m \in I \iff \exists s : \delta_{i_m}^{(s)} \neq 0$. The coordinates of the landmarks on the outer contour are calculated as $\tilde{x}_{i_m} = x_{i_m} + \delta_{i_m} \sin\theta$ and $\tilde{y}_{i_m} = x_{i_m} - \delta_{i_m} \cos\theta$. The concatenated shape vectors

$$\tilde{\mathbf{x}} = (x_1, y_1, \ldots, x_N, y_N, \tilde{x}_{i_1}, \tilde{y}_{i_1}, \ldots, \tilde{x}_{i_M}, \tilde{y}_{i_M})^\top$$

are then aligned as in the standard model using Procrustes analysis.

Principal components analysis can be applied to the shape $3N$-vectors $\mathbf{x}$ of Equation (4) analogously to standard ASM (Equations (1)–(3)). Furthermore, this model is well defined in the sense that any shape $\mathbf{x}$ generated by Equation (2) can have a non-zero separation δ_n at landmark n only if the contours are separated at this landmark in at least one training example. When all separations are zero for all training examples, a standard (single contour) ASM is recovered.

4.1 Local Appearance Models and Search

The local appearance can be modelled and searched as in the standard ASM at landmarks where no double contour is possible. At landmarks with non-zero separation, it makes sense to use profiles in the direction of the dominant axis instead of perpendicular to the contour. This is because these landmark points are constrained to move in this direction during search. Adopting another direction would necessitate a complicated and numerically unstable recalculation of the displacement at each search step. The dominant axis direction is often similar to the contour normal direction so the resulting appearance models are similar.

Possible approaches are to use (i) long fixed-length profiles which cover the corresponding landmarks on both contours, (ii) variable length profiles which cover the corresponding landmarks on both contours plus a fixed number of pixels to either side of the contour, or (iii) two separate, shorter profiles centred at the inner and outer contour landmarks. A disadvantage of the first two approaches is that the profiles are longer, requiring more training examples. The third approach was adopted here. A drawback of this approach is that the local appearance models at the inner and outer contour are treated as independent when in fact they are likely to be quite strongly coupled. (Note that a related limitation applies to standard ASM models which model the appearance of adjacent landmark independently).

When the number of training examples is limited, appearance models learned separately for each landmark can become unreliable. A windowing method is therefore adopted in which training profiles from nearby landmarks are pooled in order to estimate the appearance model. More specifically, for each landmark, profiles from the W adjacent landmarks to its left and the W landmarks to its right on the contour are used in addition to profiles at the landmark itself in order to estimate the local appearance model. This windowing is used for all landmarks (single and double contour).

Behiels *et al.* [1] reported significantly better segmentation of bones in x-ray images using alternative features. Therefore, several different features were compared here, namely raw intensity, unnormalised gradient, normalised intensity, normalised gradient, scaled intensity and scaled gradient[1].

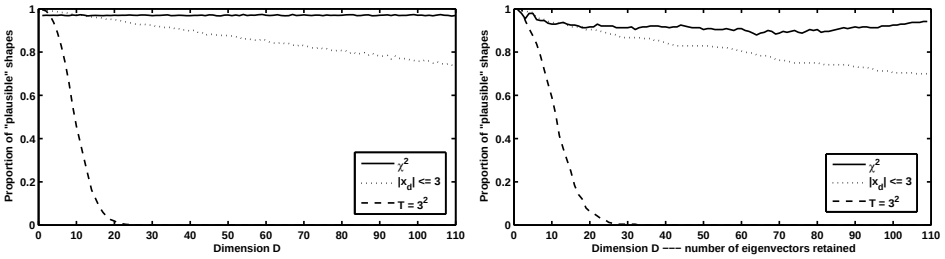
In standard ASM, Mahalanobis distance is used for measuring the quality of fit of a new profile to the learned ones. This does not take into account information about the appearance distributions off the contour. Furthermore, the underlying assumption that the profiles can be modeled as Gaussian is often not well satisfied. De Bruijne *et al.* [7] used a K-nearest neighbour classifier constructed using examples of profiles both on and off the contour to estimate the probability that a given profile lies on a contour. Note that such an estimate can only take $K+1$ different values. In this paper, distance weighted $K-NN$ is used instead. For every landmark, *on* contour profile examples are sampled as for standard ASM. In addition, *off* contour examples are obtained by sampling profiles translated in the profile direction.

The distance between two profiles $\mathbf{p}_1 = (p_1^{(1)}, \ldots, p_J^{(1)})$ and $\mathbf{p}_2 = (p_1^{(2)}, \ldots, p_J^{(2)})$ is taken to be the sum of absolute differences: $d(\mathbf{p}_1, \mathbf{p}_2) = \sum_{j=1}^{J} |p_j^{(1)} - p_j^{(2)}|$. The goodness of fit of a new profile $\mathbf{p}_f$ whose K nearest neighbours are $\mathbf{p}_1, \ldots, \mathbf{p}_K$ is defined as

$$f(\mathbf{p}_f) = \sum_{k=1}^{K} w_k \text{ where } w_k = \begin{cases} 0 & \text{if } \mathbf{p}_k \text{ is an off example} \\ \frac{1}{d(\mathbf{p}_f, \mathbf{p}_k)^2} & \text{if } \mathbf{p}_k \text{ is an on example} \end{cases} \tag{5}$$

In the unlikely event of an *on* example exactly matching $\mathbf{p}_f$, the goodness of fit is taken to be maximal. Multi-resolution ASM search is used along with the modifications needed to accommodate the double contour model.

[1] For $\mathbf{g} = (g_1, \ldots, g_M)$, the normalised vector is $\hat{\mathbf{g}} = \left(\frac{g_1}{\sum_{m=1}^{M} g_m}, \ldots, \frac{g_M}{\sum_{m=1}^{M} g_m} \right)$ and the scaled vector is $\tilde{\mathbf{g}} = \left(\frac{g_1 - \min \mathbf{g}}{\max \mathbf{g} - \min \mathbf{g}}, \ldots, \frac{g_M - \min \mathbf{g}}{\max \mathbf{g} - \min \mathbf{g}} \right)$



(a) (b)

Figure 5: The effect of dimensionality, D, on the proportion of plausible shapes in the training set using (a) samples from a Gaussian distribution with diagonal covariance matrix (10,000 examples), and (b) annotated face shapes (240 examples).

5 Shape Plausibility

PCA shape models are used to define a set of *plausible* shapes and generated shapes are constrained to membership of this set. The definition of this set is therefore important for performance. In the absence of an informative prior over the distribution of implausible shapes, plausible shapes can be defined as those with parameters $\mathbf{b}$, such that $p(\mathbf{b}) \geq p_t$, where $p(\mathbf{b})$ is a probability density function estimated from the training set and p_t is a suitable threshold chosen so that some fixed proportion of the data, γ, will exceed it [5]. Assuming a Gaussian distribution, the required decision boundary is a hyperellipsoid, and the decision rule becomes

$$\sum_{d=1}^{D} \frac{b_d^2}{\lambda_d} \leq T \tag{6}$$

where T is a threshold. This threshold limits the Mahalanobis distance and should vary with the dimensionality, so setting T to a constant is clearly inappropriate (see Figure 5). Given independent b_d parameters, the probability of plausibility using the popular univariate limits method is

$$P(|b_1| \leq B\sqrt{\lambda_1}, \ldots, |b_D| \leq B\sqrt{\lambda_D}) = (P(|b_1| \leq B\sqrt{\lambda_1}))^D \tag{7}$$

The univariate probability under assumed Gaussianity is readily available, e.g. $P(|b_d| \leq 2\sqrt{\lambda_d}) \approx 0.9544$. In a 40-dimensional eigenspace, the corresponding probability of plausibility becomes just 0.15 and this test for plausibility is clearly no longer appropriate.

The more principled method of constraining $\mathbf{b}$ to a hyperellipsoid is, to the authors knowledge, rarely used. The squared sum of independent standard Gaussian distributed variables is χ^2 distributed. Therefore, $\sum_{d=1}^{D} \frac{b_d^2}{\lambda_d}$ is χ^2 distributed with D degrees of freedom. The threshold T can therefore be chosen as the γ-quantile of this χ^2 distribution in order to have a proportion γ of plausible shapes in the training set. This quantile can be computed numerically. Figure 5 illustrates that using this method the proportion of shapes that are plausible is approximately invariant to changes in dimensionality even for a real-world, non-Gaussian shape distribution.

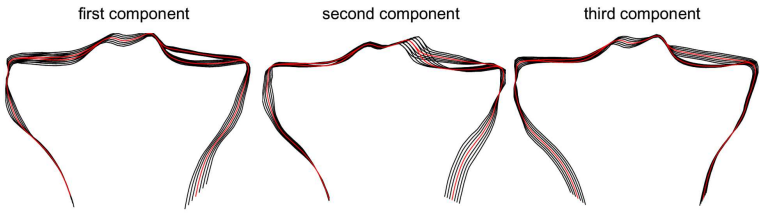
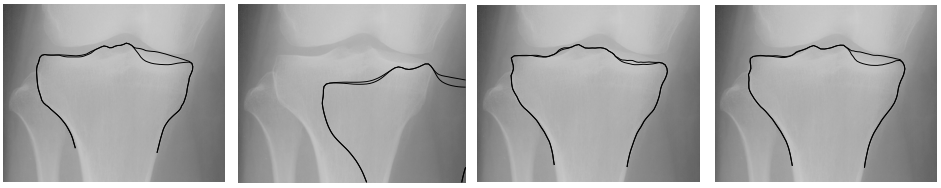


Figure 6: Tibia DCASM modes of variation for the largest 3 eigenvalues.



(a) Manual annotation (b) Initialisation(cropped) (c) Segmentation using Mahalanobis distance (d) Segmentation using weighted $K-$NN

Figure 7: DCASM tibia segmentation ($W = 10$, scaled gradient, $K = 10$).

6 Empirical Evaluation

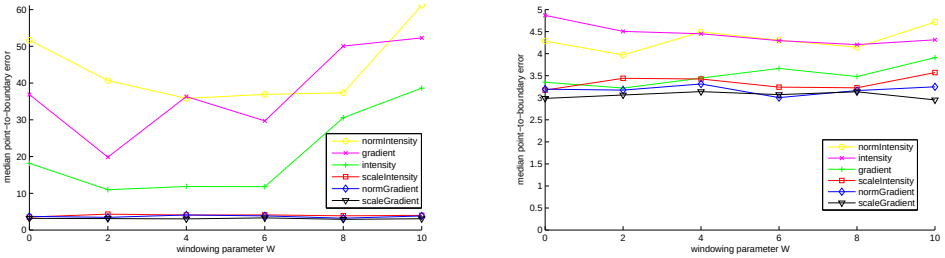
The methods described were applied to a data set of 30 standard clinical x-rays of the knee. Images of left knees were mirrored so that they appeared as right knees. All images were manually annotated and leave-one-out validation was used for evaluation. A highly approximate, manual initialisation was provided for each image. Different images showed different portions of the shaft of the tibia. The inner contours were used as reference contours. Training sets were appropriately truncated (i.e. their endpoints determined) and brought into correspondence using the MDL approach [6] with added curvature [11]. Each DCASM estimated was used to perform segmentation of its 'left out' test image.

A commonly used measurement for contour segmentation accuracy is the point-to-contour error defined as the average Euclidean distance from the obtained landmark positions to the annotated contour (which is treated as ground truth). Let E_s denote this error for the s^{th} test example. Overall performance was characterised using the median test error, $\overline{E} = \mathbf{med}_s E_s$. This gives a more meaningful indication of accuracy than the mean due to occasional gross segmentation failure.

The resulting modes of variations for the tibia are shown in Figure 6 for the largest three eigenvalues. A typical overall segmentation result with initialisation is shown in Figure 7 for the Mahalanobis distance as well as weighted $K-$NN.

Figure 8 plots the segmentation errors obtained using Mahalanobis distance and weighted $K-$NN with $K = 10$. Plots are given for window parameter values of $W = \{0,2,4,6,8,10\}$ for each of the 6 feature types. The overall lowest median error of 2.95 pixels (min $= 1.98$, max $= 25.07$) was achieved using weighted K-NN with $K = 10$, $W = 10$ and scaled gradient features (see Figure 9). The distribution of the segmentation results for this case is shown in Figure 10.

Feature type had a significant influence on accuracy. Weighted K-NN achieved better accuracy than Mahalanobis distance although the difference was small for the best feature types. The windowing parameter W and the K parameter had relatively little effect on accuracy.



(a) (b)

Figure 8: Effect of profile feature and window size on tibia median error. (a) Mahalanobis distance. (b) Weighted $K-NN$ ($K=10$). (Note differing ordinate scales).

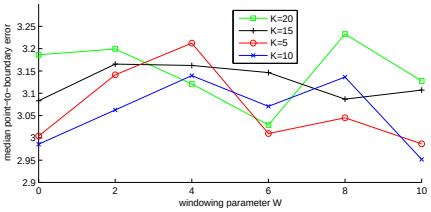


Figure 9: Effect of K and window size on median error using scaled gradient.

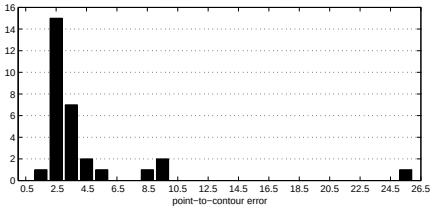


Figure 10: Histogram of segmentation errors E_s when using $K-NN$ regression ($K=10$, scaled gradient, $W=10$).

The results reported above were obtained using $D=11$ eigenvectors and setting the shape constraint parameter as $B=3$ in Equation (7). The χ^2 distribution was not initially used to set the shape constraint since the experimental results in Figure 5 suggested that the benefit might not be significant at this dimensionality. An experiment was subsequently performed using weighted $K-$NN ($K=10$, $W=6$ and scaled gradient features). When χ^2 was used to set the shape constraint such that $T=21.3$ ($\gamma=0.97=0.9973^{11}$), the median error dropped from 3.07 pixels (min $=2.08$,max $=37.60$) to 2.87 pixels (min $=1.86$,max $=51.91$). Preliminary experiments suggest that this improvement is similar in magnitude for other reasonable parameter settings.

7 Conclusions

The DCASM was introduced as a method for modelling and segmenting contours with inconsistent loops and bifurcations. Its performance was evaluated on the task of segmenting tibia contours in knee x-rays. The method should be more broadly applicable since occlusions, rotations in depth of non-convex objects in optical images, and projections of non-convex objects in transmissive imaging modalities often result in such contours.

The finding that scaled gradient features were most effective and that normalised intensity was least effective using Mahalanobis distance stands in contrast to Ref. [1] in which scaled intensity seemed to perform best for segmenting the *cranial* end of the femur. This demonstrates that even for similar applications, the optimal appearance models can vary and be difficult to find. The use of weighted K-NN resulted in far less sensitiv-

ity to feature type. Future work is planned to develop improved appearance models and search methods for the knee osteoarthritis application.

Acknowledgments

M. Seise was partially funded by EPSRC. H. H. Thodberg made available code for the MDL method. Dr B. Oliver provided knee images.

References

[1] G. Behiels, F. Maes, D. Vandermeulen, and P. Suetens. Evaluation of image features and search strategies for segmentation of bone structures in radiographs using active shape models. *Medical Image Analysis*, 6(1):47 – 62, Mar 2002.

[2] J. C. Buckland-Wright, D. G. Macfarlane, S. A. Williams, and R. J. Ward. Accuracy and precision of joint space width measurements in standard and macroradiographs of osteoarthritic knees. *Ann Rheum Dis*, 54(11):872 – 880, Nov 1995.

[3] T. F. Cootes, G. Edwards, and C. J. Taylor. Comparing active shape models with active appearance models. In *British Machine Vision Conference*, volume 1, pages 173 – 183, Sep 1999.

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

[5] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, 2004.

[6] R. H. Davies, C. Twining, T. F. Cootes, and C. J. Taylor. A minimum description length approach to statistical shape modelling. *IEEE Transactions on Medical Imaging*, 21(5):525 – 537, May 2002.

[7] M. de Bruijne, B. van Ginneken, W. J. Niessen, and M. A. Viergever. Active shape model segmentation using a non-linear appearance model: application to 3D AAA segmentation. University of Utrecht, 2003.

[8] T. E. McAlindon, C. Cooper, J. R. Kirwan, and P. A. Dieppe. Knee pain and disability in the community. *Br J Rheumatol*, 31(3):189 – 192, Mar 1992.

[9] S. Romdhani, S. Gong, and A. Psarrou. A multi-view nonlinear active shape model using kernel PCA. In *British Machine Vision Conference*, pages 483 – 492, 1999.

[10] P. D. Sozou, T. F. Cootes, C. J. Taylor, and E. C. Di Mauro. Non-linear point distribution modelling using a multi-layer perceptron. In *British Machine Vision Conference*, volume 1, pages 107 – 116, 1995.

[11] H. H. Thodberg and H. Olafsdottir. Adding curvature to minimum description length shape models. In *British Machine Vision Conference*, pages 251 – 260, 2003.

[12] B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. Romeny, and M. A. Viergever. Active shape model segmentation with optimal features. *IEEE Transactions on Medical Imaging*, 21(8):924 – 933, Aug 2002.