Learning Active Shape Models for Bifurcating Contours

Matthias Seise, Stephen J. McKenna*, Ian W. Ricketts and Carlos A. Wigderowitz

***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. Automatic segmentation of tibial and femoral contours in knee x-ray images is investigated as a step towards reliable, quantitative radiographic analysis of osteoarthritis for diagnosis and assessment of progression. Results are presented using various features, Mahalanobis distance, distance weighted $K-$nearest neighbours and two relevance vector machine based methods as quality of fit measure.**

I. Introduction

Statistical models of shape based on point distributions have enjoyed considerable success, particularly for segmentation, tracking and recognition of biological shape variation, e.g. facial and medical image analysis. The original active shape model formulation [1] 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 [2].

Consider the image contours annotated in Fig. 1. Shown are four examples from a radiographic image analysis application and four from a lip-reading 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 multiple contours, each sharing endpoints but taking different paths around the loops. Modelling these contours independently results in a poorly constrained search in which 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, contours would have differing landmarks on sections where the contours are collinear.

This work was partially supported by the EPSRC.

M. Seise, S. J. McKenna* and I. W. Ricketts are with the Division of Applied Computing, University of Dundee, DD1 4HN, Dundee, UK. e-mail:stephen@computing.dundee.ac.uk

C. A. Wigderowitz is with the Division of Orthopaedic & Trauma Surgery, University of Dundee, DD1 4HN, Dundee, UK.

In this paper, the *bifurcating contour active shape model* (BCASM) is proposed. The contours are parameterised in terms of a primary contour's landmarks along with a suitably constrained warp to a secondary 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 bifurcations 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 different methods for local appearance matching. A distance weighted K-nearest neighbour (K-NN) method is proposed and this is compared with the commonly used Mahalanobis distance. Furthermore, partially motivated by Williams *et al.* [3], the use of relevance vector machines (RVM) [4] for driving the active search is investigated.

The BCASM is extensively 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). Results are also presented for segmenting the femoral contours. Some background on this application is provided in Section II-A. Lip tracking is presented as another example application for BCASMs.

A somewhat related approach was used by Jacob *et al.* [5] to track the myocardial borders in ultrasound sequences. The inner contour (endocardium) was tracked using a dynamic contour tracking method. The search for the outer contour (epicardium) was constrained by using a difference-shape-space learned from the difference of the manually annotated contours. An assumption was that the endocardium was easier to track and as such guided the search for the epicardium. This assumption is not necessary in the BCASM; rather contours are treated equally by the appearance model and the search.

Another related approach was proposed by Luettin *et al.* [6] for tracking lips using ASMs. The inner and outer contours of the lips were manually annotated and normalised by scaling, rotating and translating so that the line connecting the endpoints of the lips was horizontal with unit length and centred on the point of origin. The landmarks were set evenly spaced along this horizontal. Thus, it was possible to build the shape model by regarding only the vertical parameters. In subsequent work, they discarded this shape model in favour of a standard point distribution model since it did not fully capture the large variability of the inner contour of the lips [7]. Furthermore, equally spaced landmark points gave only imprecise correspondence and therefore a weaker shape model. This problem is avoided here by using an automatic method based on minimum description length criteria [2].

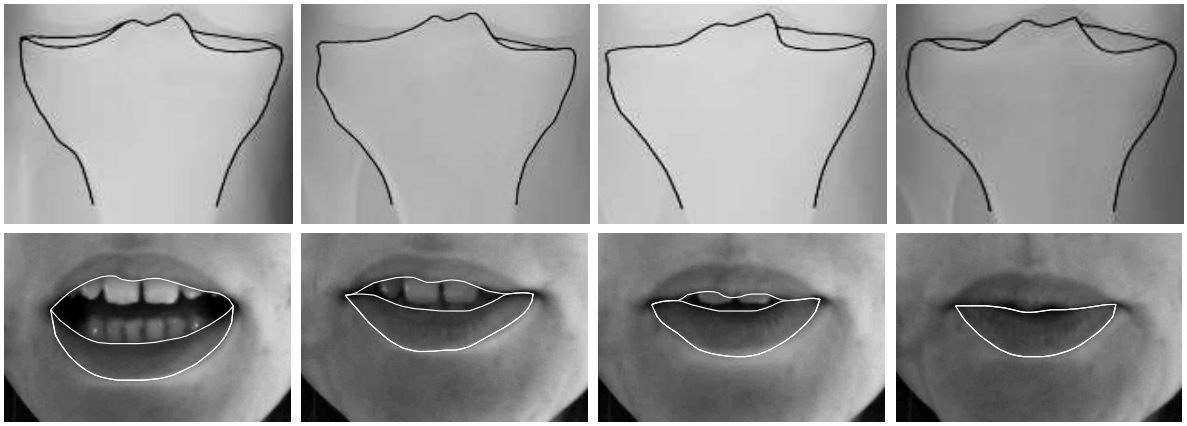


Fig. 1. Shape examples annotated with looping contours. Top: radiographs of the tibia. Bottom: a lip-reading application. Note that the number of loops and their positions vary.

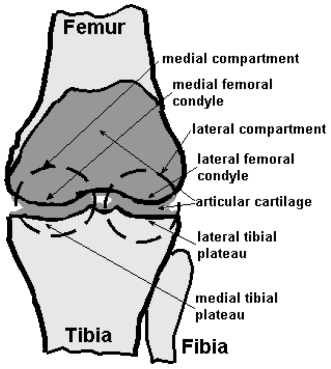


Fig. 2. Anatomy of the knee.

Fig. 3. Clinical x-ray (enhanced).

II. Background

The BCASM was initially developed to segment the knee joint from standard x-ray images as a first step towards the automatic assessment of osteoarthritis (OA). This section gives a brief overview of the medical background to OA and also briefly reviews and describes the standard ASM.

A. Osteoarthritis

Osteoarthritis is the most common joint disease and the most common cause of disability in older people [8], resulting in significant economic costs to 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 Fig. 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. Some of these signs are difficult to quantify but for clinical trials, classification or grading of OA progression is often needed.

Unfortunately, there is still no accepted unifying standard for the assessment of OA. A review of existing assessment methods, such as the first by Kellgren and Lawrence [9], the most recent and current standard used by the Osteoarthritis Research Society developed by Altman *et al.* [10], and a new proposal, is given by Nagaosa *et al.* [11]. Most of these methods share the simple approach of visual comparison to standard radiographs is used, mainly based on joint space width and osteophytes. This leads to poor repeatability [12]. In particular, the *minimum* joint space width (JSW), normally used by physicians as a quantitative measurement of joint space, results in large inter- and intra-observer variation [13]. In structure-modifying drug trials, the change of the joint space over time, is often used as an output measure. Therefore, it is mandatory to assess it with high precision and reproducibility. The exact segmentation of the tibial plateaux (as presented here) and of the femoral condyles (presented in earlier work [14] and partially here) is an important step towards this goal.

B. Active Shape Models

Before introducing the BCASM, the standard ASM is first briefly reviewed. Introduced originally by Cootes and Taylor [1], ASMs are widely used, especially in medical image segmentation. Many extensions have been developed, from non-linear shape models based on Gaussian mixture models to wavelet based shape models as well as improved models of local appearance and search techniques. Applications and extensions are so numerous that they cannot be reviewed here in detail. Only a few will be mentioned. A good overview is given by Cootes and Taylor [15].

1) Theory: Given a training set of S images in which the objects of interest are suitably annotated, statistical shape and appearance models can be estimated [15]. 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 [2]. 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:

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

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

$$\mathbf{x}_s \approx \overline{\mathbf{x}} + \Phi\mathbf{b}_s \ , \quad (2)$$

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

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

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 [15] 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 $3\sqrt{\lambda_d}$. The 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 appearance models learned for each resolution. The segmentation result at each resolution is used to initialise the search at the next resolution.

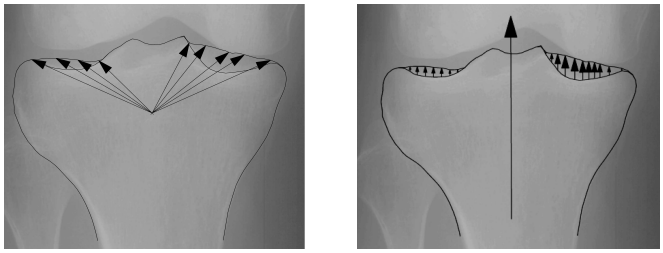
2) Review: Several improvements to the standard ASM have been proposed. For example, more complex features characterising texture have been used for appearance modelling [16]. Active appearance models which model 2D appearance as well as shape variation using PCA are useful in applications such as human face analysis [17] although they often result in lower accuracy localisation of contours than ASM [18]. When a linear model of shape variation is inadequate, non-linear models have been used, e.g. [19], [20].

Since their conception, ASMs have been used for medical image analysis [21], mainly to locate structures in image modalities such as MRI and ultrasound. In Smyth *et al.* [22] ASMs were used to segment vertebral shapes from dual x-ray absorptiometry (DXA) in order to find vertebral fractures. The initial findings were good with the same accuracy for the automatic method as for manual annotation. DXA radiographs were also used by Sotoca *et al.* [23] and Thodberg and Rosholm [24] to segment parts of the metacarpals, whole metacarpals, medial phalanxes and proximal phalanxes. Thodberg and Rosholm [24] extended the ASM to the More Active Shape Model (MASM) using a 'translation operator' to provide the active search with methods to move the shape orthogonal to the local search directions. This extension was necessary since only a part of the shaft on the metacarpal was segmented. However, if entire bone contours are segmented, the standard ASM can be used [23]. Quantitative segmentation errors were not reported in either paper but the segmentation failure rate was said to be small. Standard ASMs were used by Hutton *et al.* [25] to locate landmark points on standard x-rays of the head but with unsatisfactory results. Vogelsang *et al.* [26] and Kohonen *et al.* [27], [28] used a combination of the ASM shape model, edge-based energy term and simulated annealing as a search algorithm to segment hand radiographs, and lumbar x-ray images. Zamora *et al.* [29] used a combination of template matching, ASM and deformable models to segment the vertebrae from x-ray images of the head.

Probably the most closely related work was published by Behiels *et al.* [30]. They compared different search strategies (standard ASM, a minimal-cost-path (MCP) extension to ASM, and a MAP estimate for simultaneous optimisation of shape and pose parameters) and different feature types for the segmentation of different bones. The presented results indicated that MCP performed best in all cases. However, the results must be handled with care since the initialisation was almost perfect (using the manually annotated contours as initialisation) and no multi-resolution search was used. Nevertheless, the results give an idea of possible strategies to minimise segmentation errors.

III. ASM for Bifurcating Contours

In order to model bifurcating contours such as those in Fig. 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 *primary contour* and the bifurcating contour is represented in terms of landmarks on this primary contour along with a constrained warp to the secondary contour. The warp defines corresponding landmarks on the secondary 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 primary contour $(\overline{x}, \overline{y}) = \left(\frac{1}{N}\sum_{n=1}^{N}x_n, \frac{1}{N}\sum_{n=1}^{N}y_n\right)$ and the corresponding landmark (x_n, y_n) (see Fig. 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 Fig. 1, a suitable warp can be



(a) (b)

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

achieved by parallel displacement of the landmarks, i.e. by translating each of the N landmarks on the primary contour, (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 Fig. 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 was computed. 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. For the lip tracking application, θ was set orthogonal to the line connecting the endpoints of the lower lip.

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

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

The correspondence between the primary contour and the secondary 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 primary contours of each shape now need to be brought into correspondence. This can be achieved using the minimum description length method [2].

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 primary contour landmarks and those secondary 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 secondary 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.

As explained in detail in [31] there are different methods to define plausible shapes. In most ASM applications, the D-dimensional shape parameters $\mathbf{b}$ from Equation (2) are constrained by $|b_d| < 3\sqrt{\lambda_d}$ for all $d = 1, ..., D$. Here, the more principled method was adopted of constraining $\mathbf{b}$ to a hyperellipsoid by

$$\sum_{d=1}^{D} \frac{b_d^2}{\lambda_d} < T \quad (5)$$

with T being the γ-quantile of the χ^2 distribution in order to have a proportion γ of plausible shapes in the training set. This quantile can be computed numerically and $\gamma = 0.98$ is used in all experiments reported here.

A. 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 warp 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 second approach would require the comparison of profiles with different lengths. 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 landmarks 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. At both ends of the contour the training profiles of the end and adjacent $W-1$ landmarks are used. This windowing is used for all landmarks (both single and double contour).

Behiels *et al.* [30] reported significantly better segmentation of the femur 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 (as used by Behiels *et al.*), and additionally G-scaled intensity and G-scaled gradient. For an arbitrary vector $\mathbf{g} = (g_1, \ldots, g_M)$, the normalised vector is

$$\hat{\mathbf{g}} = \left(\frac{g_1}{\sum_{m=1}^{M} g_m}, \ldots, \frac{g_M}{\sum_{m=1}^{M} g_m} \right) ,$$

the scaled vector is

$$\tilde{\mathbf{g}} = \left(\frac{g_1 - \min_m g_m}{\max_m g_m - \min_m g_m}, \ldots, \frac{g_M - \min_m g_m}{\max_m g_m - \min_m g_m} \right)$$

and the G-scaled vector is defined as

$$\tilde{\tilde{\mathbf{g}}} = \left(\frac{g_1 - \overline{\mathbf{g}}}{\sigma_{\mathbf{g}}}, \ldots, \frac{g_M - \overline{\mathbf{g}}}{\sigma_{\mathbf{g}}} \right) ,$$

with $\overline{\mathbf{g}}$ being the mean and $\sigma_{\mathbf{g}}$ the standard deviation of the vector $\mathbf{g}$.

In standard ASM, Mahalanobis distance is used for measuring the quality of fit of a new profile to the learned model. This does not take into account information about the appearance distributions off the contour. Furthermore, the underlying assumption that the profiles can be modelled as Gaussian is often not well satisfied. De Bruijne *et al.* [32] 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

$$f(\mathbf{p}_f) = \sum_{k=1}^{K} w_k$$
$$\text{where } w_k = \begin{cases} 0 & \text{if } \mathbf{p}_k \text{ is an off example} \\ \frac{1}{d(\mathbf{p}_f, \mathbf{p}_k)^2} & \text{if } \mathbf{p}_k \text{ is an on example} \end{cases} \quad (6)$$

In the unlikely event of an *on* example exactly matching $\mathbf{p}_f$, the goodness of fit is taken to be maximal.

In standard ASM, multi-resolution search is used to avoid convergence to the wrong local structure and to speed up the search. Usually the search only changes the resolution of the underlying image, i.e. by subsampling the appearance model. Here a further step is used by subsampling the shape space too, as originally proposed in [33]. A shape model is built as described above for all L levels of search, at the highest resolution using all N landmarks, and at each lower level using every second landmark from the level above. This not only speeds up the search, it also resolves some problematic issues when the resolution of the image is so low compared to the shape that different landmarks have the same coordinates and the perpendicular direction is not well defined. This kind of multi-resolution search is of course only useful if the total number of landmarks fulfils $N \gg 2^L$.

B. Relevance Vector Machines for Learning Displacements

Williams *et al.* [3] used RVMs to build displacement experts and drive a tracking algorithm trained on image patches. It should also be possible to use RVM's to improve the appearance model an obtain more robust and accurate ASM search. Explaining RVM in detail is out of the scope of this paper; see for instance [4]. (Note that the notation in the following brief explanation is taken from [4].) In brief, RVMs are a Bayesian treatment of sparse learning and are used here for regression. The output functional is

$$y(\mathbf{x}; \mathbf{w}) = \sum_{n=1}^{N} w_n K(\mathbf{x}, \mathbf{x}_n) + w_0 = \mathbf{w}\boldsymbol{\phi}(\mathbf{x})$$

where $\mathbf{x}$ is an input vector, $y : \mathbb{R}^M \to \mathbb{R}$ the output function, K is the kernel function and $\mathbf{w} = [w_0, \ldots, w_N]^\top$ the weights determined by training. It is assumed that the training data are sampled with additive noise

$$t_n = y(\mathbf{x}_n; \mathbf{w}) + \epsilon_n$$

where $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$. The prior over the weights is a zero-mean Gaussian with diagonal covariance matrix. The use of such a prior enables sparsity to be obtained. After determining the weights from the training data, prediction of the target value t_* for a new unknown datum $\mathbf{x}_*$ results in

$$t_* \sim \mathcal{N}(y_*, \sigma_*^2) \quad (7)$$

with

$$y_* = \sum_{n=1}^{N} \mu_n K(\mathbf{x}_*, \mathbf{x}_n) = \boldsymbol{\mu}^\top \boldsymbol{\phi}(\mathbf{x}_*)$$
$$\sigma_*^2 = \sigma_{MP}^2 + \boldsymbol{\phi}(x_*)^\top \boldsymbol{\Sigma} \boldsymbol{\phi}(x_*)$$

where $\boldsymbol{\mu}, \sigma_{MP}^2$ and $\boldsymbol{\Sigma}$ are also determined during the training steps.

The idea was to train RVMs for the appearance at every landmark and use them to drive active search using the probabilistic outputs. In detail, $2T+1$ fixed length profiles $\mathbf{p}_1, \ldots, \mathbf{p}_{2T+1}$ were sampled orthogonal to the contour, the centre offset by $t_1 = -T, \ldots, t_{2T+1} = T$ pixels from the contour. These t_i were also the target values for training the RVMs. The RVMs were trained for the current resolution and landmark to output the expected offset from the true contour for any input profile. For training, the source code provided by M Tipping based on Ref. [34] was used.

Two different methods are proposed here to use trained RVMs during active search. The first method is slow but should yield superior results since it uses the probabilistic prediction of the target value from Equation (7). In detail, for

every landmark, search profiles $\mathbf{p}_1, \ldots, \mathbf{p}_M$ were sampled orthogonal to the contour($M \leq 2T+1$). The output of the RVM for a profile $\mathbf{p}_m$ is a Gaussian posterior density over t_m with the mean $\overline{t_m}$ and the standard deviation σ_{t_m} (Equation (7)). Since the error-bars have the strange property of being smallest when the input data are far from the training set (see [4], App. D), σ_{t_m} was set to infinity and $\overline{t_m}$ to zero if $|\overline{t_m}| > T$. Assuming conditional independence, the landmark position was updated to t_* which minimises $f(t) = \sum_{m=0}^{M} \frac{(t-\overline{t_m})^2}{\sigma^2_{t_m}}$. Using all the information around the landmark should lead to more robust fitting and therefore higher accuracy.

In the second method, for every landmark only one profile $\mathbf{p}$ centred on and orthogonal to the contour is sampled. The output of the RVM gives the expected displacement of the current contour to the 'real' contour. The current landmark is updated by the expected offset $\bar{t}$. This method is much faster, since for every landmark the RVM is called once instead of M times. The performance of both methods suffers from the long training times of RVMs, but this is done offline. The expected sparsity should give a performance gain in terms of computational costs for the second method compared to the K-NN method and be comparable to the use of Mahalanobis distance.

IV. Empirical Evaluation

A. Bone segmentation

The methods described were applied to a data set of 30 standard clinical x-rays of non-osteoarthritic knees. Concavities in the tibial plateaux result in distinct image contours corresponding to the anterior and posterior rims of the plateaux. It is very difficult to determine which contour is which on the basis of the AP radiograph. The contours are therefore referred to as the *inner* contour and the *outer* contour in a 2D sense. These double contours are not always present on both the lateral and medial plateaux. Furthermore, the contour bifurcation points vary quite widely between example images.

The radiographs were digitised with a resolution of 150dpi (1 pixel is about 0.17mm) and 8 bit grayscale depth. The image size was between 1312x928 and 1760x1408 pixels. Images of left knees were mirrored so that they appeared as right knees. Inner and outer contours were manually annotated in all images and leave-one-out validation was used for evaluation. Automatic initialisation was based on a simple threshold method to find the centre line of the shafts of the tibia and femur and the gap between the heads of the bones. Additionally, average rotation and scaling of the initialisation template were learned from the training data. The inner contours were used as primary contours and the outer contours as secondary contours. Different images showed different portions of the shaft of the tibia. Therefore, the MDL approach [2] with added curvature [35] was used in two steps. The first step truncated the training set, i.e. the endpoints were loosely determined. In the second step, the factor for the curvature term was increased to bring the shapes into better correspondence by taking advantage of the specific shape around the tibial plateaux. Each BCASM estimated was used to perform segmentation of its 'left out' test image.

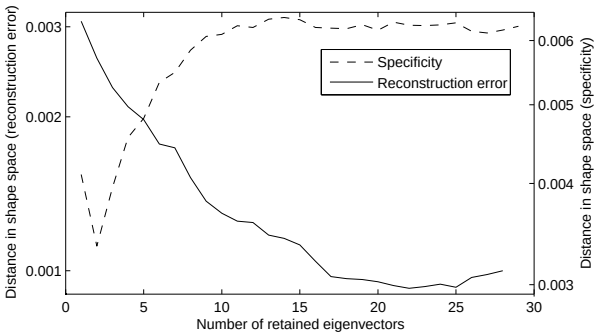


Fig. 6. Generalisation and Specificity of the Tibia measured in the shape space (note the different logarithmic scale).

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 mean test error, $\overline{E}$ and its standard deviation σ_E.

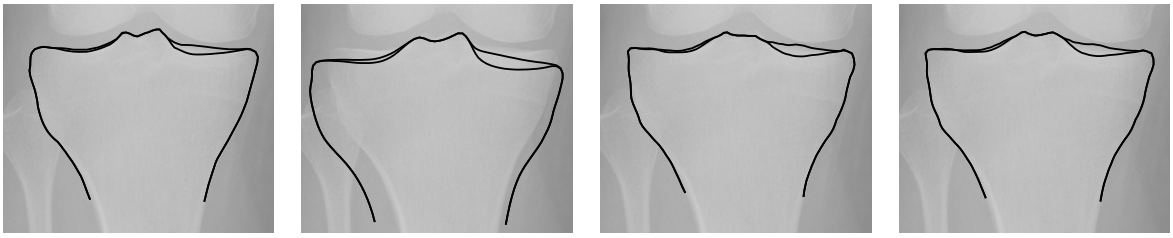
There are different approaches possible to find the appropriate number, D, of retained dominant eigenvectors after PCA. The commonly used one was described in Section II-B, using a fixed proportion of the total variance. A different heuristic also described in [15] uses generalisation ability and specificity instead. The generalisation ability is calculated using leave-one-out reconstruction by building the shape model from the remaining training examples and calculating the mean point-to-point reconstruction error in the *shape space*. The generalisation of the model is then the mean over all leave-one-out experiments. The specificity was calculated as described in [36] by sampling randomly 1000 plausible shapes, calculating the point-to-point errors for all training shapes per sample, and summing their minimum values. Reconstruction errors and specificity measure could also be computed in image space, resulting in error measures in pixels. Since only the shapes of the curves are of interest, this is not necessary.

1) Tibia segmentation: Fig. 6 shows specificity and reconstruction error curves for the tibia. Considering both these curves, $D = 17$ was chosen.

Automatic initialisation failed completely in 1 of the 30 examples so that the point-to-contour errors were larger than 20 pixels for all tested parameter combinations. Therefore, this example was excluded in the following quantitative evaluations.

A typical overall segmentation result with automatic initialisation is shown in Fig. 5 for the Mahalanobis distance as well as weighted $K-$NN. The BCASM algorithm was evaluated extensively for all 8 features, window parameter values of $W = \{0, 2, 4, 6, 8, 10\}$, Mahalanobis distance and weighted $K-$NN with $K = \{5, 10, 15, 20, 25, 30\}$. Fig. 7 plots the segmentation errors obtained using Mahalanobis distance and weighted $K-$NN with $K = 5$. Plots are given for window parameter values of $W = \{0, 2, 4, 6, 8, 10\}$ for each of the 8 feature types with median errors lower than 10 pixels.

The overall lowest mean error of 2.54 pixels ($\min = 1.60$,



(a) Manual annotation (b) Initialisation (c) Segmentation using Mahalanobis distance ($E_s = 1.81$) (d) Segmentation using weighted K-NN ($E_s = 1.60$)

Fig. 5. BCASM tibia segmentation ($W = 10$, scaled gradient, $K = 10$).

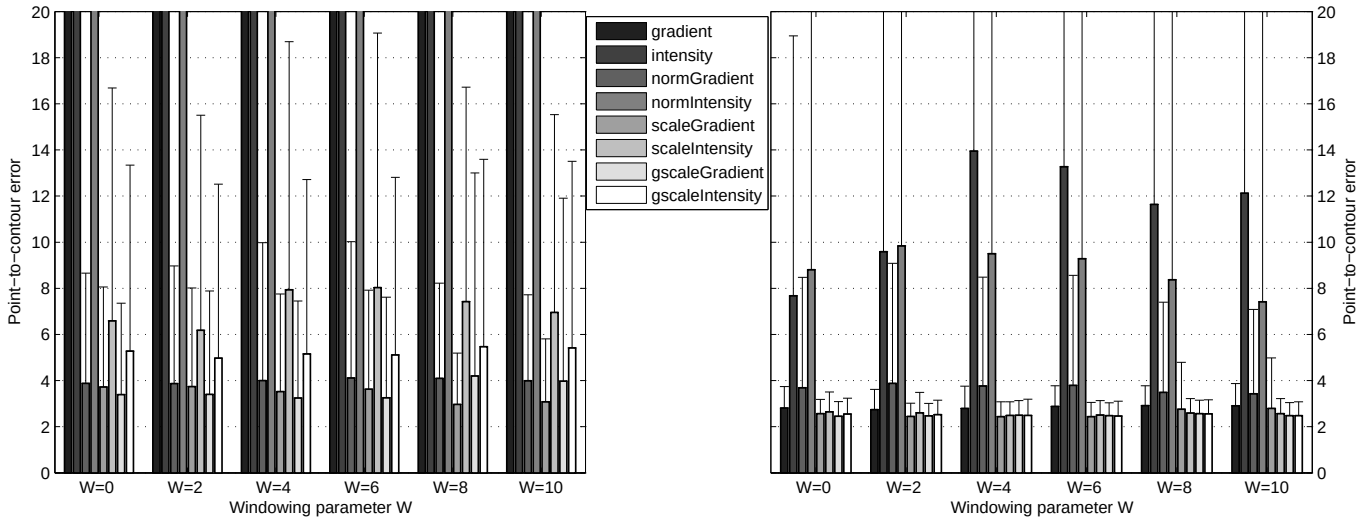


Fig. 7. Effect of profile feature and window size on $\overline{E}$ and σ_E. Left: Mahalanobis distance. Right: Weighted $K - NN$ ($K = 5$). (Note that both graphes are cropped to show the important parts of the data.)

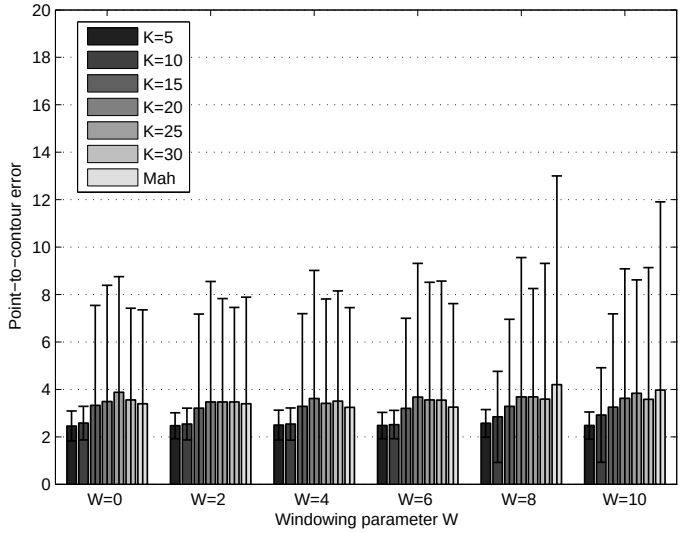


Fig. 8. Effect of K and window size on $\overline{E}$ and σ_E using G-scaled gradient.

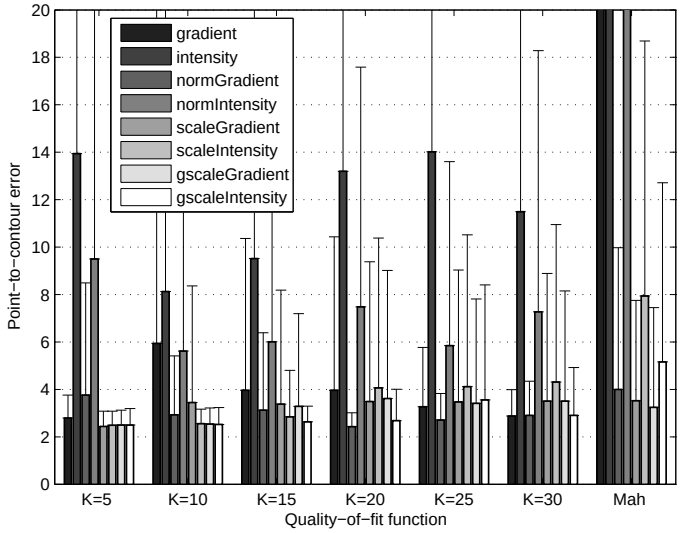


Fig. 9. Effect of K and feature type on $\overline{E}$ and σ_E using $W = 0$.

$\max = 4.37$, $\sigma_E = 0.65$) was achieved using weighted K-NN, $K = 5$, $W = 6$ and G-scaled gradient features. Depending on whether Mahalanobis distance or weighted K-NN was used, the feature types have a significant influence on accuracy. Unnormalised intensity and gradient as well as normalised intensity performed worst using Mahalanobis distance with mean errors greater than 20 pixels for all parameter settings, whereas normalised gradient and G-scaled gradient performed best. This finding stands in contrast to Ref. [30] in which scaled intensity seemed to perform best for segmenting the *cranial* end of the *femur*. When using weighted K-NN, the difference between the features types was much smaller, with even the worst performers, usually unnormalised and normalised intensity, having far less than 20 pixels mean error.

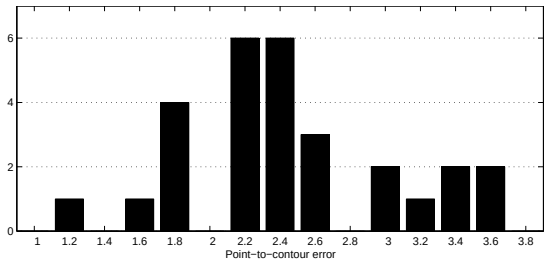


Fig. 10. Histogram of segmentation errors E_s when using $K-NN$ regression ($K = 25$, normalised gradient, $W = 2$).

Overall, the windowing parameter W had almost no influence on the accuracy independent of the other parameters (Fig. 8), whereas the K parameter had, dependent on the feature type used, an effect on the mean accuracy and standard derivation. However, the influence was significant in only a few cases (Fig. 9).

The distribution of the point-to-contour errors for one of the best results is shown in Fig. 10 using normalised gradient, $W = 2$ and $K = 25$.

2) Femur segmentation: The BCASM was also evaluated on the task of segmenting the femur and, since no multiple contours occur, it is equivalent to the standard shape model. The data set consisted of the same radiographs as for the segmentation of the tibia. Experiments were performed for the same parameter sets as in the tibia experiments. For the femur the automatic initialisation failed completely in 2 out of the 30 images, i.e. the segmentation error was larger than 20 pixels for *all* tested parameter sets. Detailed results, excluding these 2 cases, using Mahalanobis distance and K-NN ($K = 10$, no windowing) are given in Table I with normalised gradient, scaled gradient and scaled intensity as feature types. These parameter sets showed the best results. The other findings were similiar to the tibia results.

3) ASM and the Relevance Vector Machine: RVMs were trained using profiles of length 11 and offset $T = 7$ to give enough information about the appearance on and off the contour. The learning algorithm, using Gaussian kernels ($\sigma = 0.5$) centred on the test data, did not converge for all feature types. RVMs were only trained for normalised gradient, scaled gradient and scaled intensity, since these gave the best results using Mahalanobis distance and weighted K-NN. No windowing was used for the same reason.

Detailed results for the RVM segmentation compared to the same experiments using Mahalanobis distance and K-NN ($K = 10$, no windowing) are given in Table I. Weighted K-NN always performed significantly better than the first variant of using RVM. The median and the minimum segmentation errors were significantly worse using RVMs compared to weighted K-NN. The second (fast) method performed almost as well as K-NN and Mahalanobis distance. The sparsity of the RVMs depends strongly on the actual appearance around the landmark point and can range between 4 and 435 relevance vectors at the highest resolution (maximum of $435 = 29 \cdot 15$). The number of relevance vectors was about 30 for normalised gradient, 340 for scaled gradient and 100 for scaled intensity.

B. Lip tracking

As a second example application, BCASM was used for lip tracking. A sequence of 194 images (135x214 pixel, 8 bit grayscale, uncompressed) was used. The first 33 images were manually annotated as shown in Fig. 1 to provide the training set and the remainder were used as test images. Manual annotation was performed on colour images whilst the algorithm ran on grayscale images.

Annotation of the lower lip as a closed contour was used to provide the appearance model with information about the appearance of the left and right ends of the lips. (Experiments with only the inner contours of the lips failed since the search and sampling profiles were in the direction of the dominant axis, so the model could not converge to the left and right corners of the mouth.)

The primary contour was the closed contour around the lower lip and the secondary contour was the inner contour of the upper lip. The MDL approach was used to bring the primary contours into correspondence, this time without using the curvature term. The $D = 6$ dominant eigenvectors were used to build the shape model. Active search was initialised in each frame with the shape and pose of the previous frame and in the first frame with the mean shape and pose. Since multi-resolution search was used and the images were relatively small, initialisation with mean shape and pose in every image would have given qualitatively similar results.

Results are shown for normalised gradient as feature type, no windowing ($W = 0$) and Mahalanobis distance as quality-of-fit measure. The whole sequence can be downloaded from `http://www.computing.dundee.ac.uk/staff/mseise/work.html`. Results for some frames are shown in Fig. 11.

V. Discussion and Conclusions

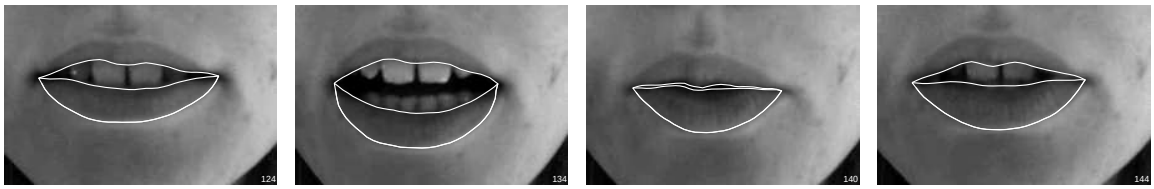
The BCASM was introduced as a method for modelling and segmenting contours with inconsistent loops and bifurcations. Its performance was mainly evaluated on the task of segmenting tibia contours in knee x-rays and secondly on a lip tracking application. The method should be more broadly applicable since occlusions, rotations in depth of non-convex objects in optical images, and projections of non-convex objects in transmissive imaging modalities often result in bifurcating contours. The proposed BCASM was explained here for shapes with two contours since in the presented applications only two contours could occur. For applications with shapes having more than two contours, the method can be extended in a straight-forward way and as such can be considered as a generalised ASM for shapes with multiple bifurcations.

The BCASM algorithm was evaluated extensively for the segmentation of the tibia. The results with mean errors between 2 and 3 pixels are promising but should be improved for practical applications like measuring the joint space. One of the limiting factors for higher accuracy seems to be the manual annotation, even for non-osteoarthritic bones. Contours can be blurred, occluded by other bones, noise and clutter. Therefore, it is not possible to manually annotate the contours of the

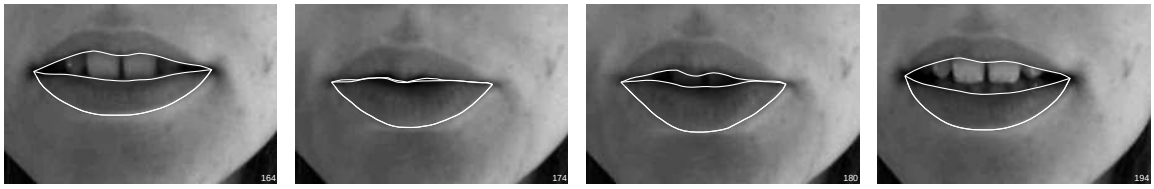
Feature	Method	Median error $\mathbf{med}E_s$	Mean error $\overline{E}$	Standard deviation σ_E	Minimal error $\min_s E_s$	Maximal error $\max_s E_s$
normalised gradient	RVM	2.00	4.71	9.18	1.26	41.43
	RVM (fast)	1.78	7.62	21.44	1.16	111.29
	Mahalanobis distance	3.22	12.87	24.01	0.86	98.34
	weighted K-NN ($K = 10$)	1.05	1.36	0.81	0.75	4.42
scaled gradient	RVM	4.63	9.27	15.16	1.76	79.23
	RVM (fast)	1.42	1.52	0.47	0.88	2.79
	Mahalanobis distance	1.01	1.22	0.41	0.82	2.60
	weighted K-NN ($K = 10$)	1.03	1.41	1.29	0.78	7.55
scaled intensity	RVM	6.98	7.83	5.15	1.65	24.59
	RVM (fast)	1.73	7.14	20.43	0.96	106.82
	Mahalanobis distance	2.73	14.10	25.58	0.99	93.37
	weighted K-NN ($K = 10$)	1.13	6.27	18.67	0.73	93.53

TABLE I

SEGMENTATION ERRORS FOR FEMUR USING RVM, MAHALANOBIS DISTANCE RESPECTIVELY WEIGHTED K-NN (SEE TEXT).



(a) Frame 124/194, $E_s = 0.76$ (b) Frame 134/194, $E_s = 0.70$ (c) Frame 140/194, $E_s = 0.88$ (d) Frame 144/194, $E_s = 1.89$



(e) Frame 164/194, $E_s = 0.59$ (f) Frame 174/194, $E_s = 0.71$ (g) Frame 180/194, $E_s = 1.23$ (h) Frame 194/194, $E_s = 0.57$

Fig. 11. Results for the lip tracking application using Mahalanobis distance and normalised gradient

bone to one pixel accuracy. Furthermore, systematic errors can occur,from the use of linear interpolation between landmark points in the MDL approach. Consequently, the shape and, more importantly, the appearance models are trained on noisy data which can lead to lower accuracy. Using the manual annotation as ground truth can of course also result in wrong error measurements, since it is conceivable that the automatic segmentation is closer to the truth than manual segmentation. Nevertheless, the reported results show that automatic segmentation of the tibia is feasible using BCASM.

A probable reason for the poor performance of the first method of using RVMs is the degenerative covariance function which leads to unusable predictive variance. This weakness of RVMs was recently highlighted by Rasmussen and Quiñonero-Candela [37]. They demonstrated significantly worse prediction rates compared to Gaussian processes which are non-sparse. The fast method using RVM shows promising results. The accuracy in the majority of the cases is as good as using K-NN or Mahalanobis distance and has the same magnitude as the expected error in the manual annotation. Future developments in sparse Bayesian inference are likely to yield improved performance.

The finding that (G-)scaled gradient features were most effective and that normalised intensity was least effective using Mahalanobis distance stands in contrast to Ref. [30] in which scaled gradient seemed to perform significantly worse than scaled intensity 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. Direct comparison to the results for segmenting the tibia is not possible since important information like image resolution and how the tibia was annotated (outer, inner or both contours) are missing. Furthermore, the active search was initialised using the known ground truth and was not multi-resolution. Using the mean shape aligned to the ground truth as initialisation, Behiels *et al.* [30] reported a mean point-to-boundary error for the tibia of 2.40 pixels using standard ASM and were 1.71 pixels using MCP. Results for segmenting the femur were also reported with a mean error of 3.39 pixels using ASM and 1.96 pixels using MCP. Table I shows some results (no windowing) for the same task using the described framework with Mahalanobis distance and weighted K-NN as quality-of-fit measures. The overall best mean point-to-contour error was given using scaled gradient, Mahalanobis distance and windowing, $W = 2$ and was 1.19 pixels, significantly smaller than reported by Behiels *et al.* [30].

The use of Mahalanobis distance as quality-of-fit measure often yields good results in terms of mean error and is much

faster than using weighted K-NN. The strong point of K-NN is the reduced sensitivity to feature type and the smaller variance, especially with small values of K.

The reported results are a promising step towards assessing osteoarthritis from standard clinical x-rays. The results show that with the proposed BCASM the segmentation of anterior *and* posterior rims of the tibial plateaux is achieved. This segmentation along with the segmentation of the femoral condyles should lead to better measurements, more highly correlated with the actual volume of cartilage than the commonly used minimum joint space width. Furthermore, a more accurate segmentation offers the potential to automatically measure and count osteophytes and ultimately to define a new outcome measure for the progression of osteoarthritis.

ACKNOWLEDGMENTS

H. H. Thodberg and M. Tipping made available the code for the MDL method and the RVM respectively. Dr B. Oliver provided knee images.

REFERENCES

[1] T. Cootes and C. Taylor, "Active shape models — 'smart snakes'," in *British Machine Vision Conference*, 1992, pp. 267–275.

[2] 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*, vol. 21, no. 5, pp. 525–537, 2002.

[3] O. Williams, A. Blake, and R. Cipolla, "Sparse Bayesian learning for efficient visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1292–1304, 2005.

[4] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[5] G. Jacob, J. A. Noble, C. Behrenbruch, A. D. Kelion, and A. P. Banning, "A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography." *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 226–238, 2002.

[6] J. Luettin, N. A. Thacker, and S. W. Beet, "Active shape models for visual speech feature extraction," in *Speechreading by Humans and Machines*, ser. NATO ASI Series, Series F: Computer and Systems Sciences, D. G. Storck and M. E. Hennecke, Eds. Berlin: Springer Verlag, 1996, vol. 150, pp. 383–390.

[7] J. Luettin and N. A. Thacker, "Speechreading using probabilistic models," *Computer Vision and Image Understanding*, vol. 65, no. 2, pp. 163–178, 1997.

[8] T. E. McAlindon, C. Cooper, J. R. Kirwan, and P. A. Dieppe, "Knee pain and disability in the community," *British Journal of Rheumatology*, vol. 31, no. 3, pp. 189–192, 1992.

[9] J. H. Kellgren and J. S. Lawrence, "Radiological assessment of osteo-arthrosis," *Annals of the Rheumatic Diseases*, vol. 16, no. 4, pp. 494–502, 1957.

[10] R. D. Altman, M. Hochberg, W. A. Murphy, F. Wolfe, and M. Lequesne, "Atlas of individual radiographic features in osteoarthritis." *Osteoarthritis and Cartilage*, vol. 3 Suppl A, pp. 3–70, 1995.

[11] Y. Nagaosa, M. Mateus, B. Hassan, P. Lanyon, and M. Doherty, "Development of a logically devised line drawing atlas for grading of knee osteoarthritis," *Annals of the Rheumatic Diseases*, vol. 59, no. 8, pp. 587–95, 2000.

[12] K. P. Günther and Y. Sun, "Reliability of radiographic assessment in hip and knee osteoarthritis," *Osteoarthritis and Cartilage*, vol. 7, no. 2, pp. 239–246,1999.

[13] 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," *Annals of the Rheumatic Diseases*, vol. 54, no. 11, pp. 872–880, 1995.

[14] M. Seise, S. J. McKenna, I. W. Ricketts, and C. A. Wigderowitz, "Segmenting tibia and femur from knee x-ray images," in *Medical Image Understanding and Analysis*, 2005, pp. 103–106.

[15] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for computer vision," University of Manchester, Tech. Rep., 2004. [Online]. Available: http://www.isbe.man.ac.uk/~bim

[16] 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*, vol. 21, no. 8, pp. 924–933, 2002.

[17] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *European Conference on Computer Vision*, vol. 2, 1998, pp. 484–498.

[18] T. F. Cootes, G. Edwards, and C. J. Taylor, "Comparing active shape models with active appearance models," in *British Machine Vision Conference*, vol. 1, 1999, pp. 173–183.

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

[20] 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*, vol. 1, 1995, pp. 107–116.

[21] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image and Vision Computing*, vol. 12, no. 6, pp. 355–365, 1994.

[22] P. P. Smyth, C. J. Taylor, and J. E. Adams, "Vertebral shape: automatic measurement with active shape models," *Radiology*, vol. 211, no. 2, pp. 571–578, 1999.

[23] J. M. Sotoca, J. M. Iesta, and M. A. Belmonte, "Hand bone segmentation in radioabsorptiometry images for computerised bone mass assessment," *Computerized Medical Imaging and Graphics*, vol. 27, no. 6, pp. 459–467, 2003.

[24] H. H. Thodberg and A. Rosholm, "Application of the active shape model in a commercial medical device for bone densitometry," in *British Machine Vision Conference*, vol. 1, 2001, pp. 43–52.

[25] T. J. Hutton, S. Cunningham, and P. Hammond, "An evaluation of active shape models for the automatic identification of cephalometric landmarks." *European Journal of Orthodontics*, vol. 22, no. 5, pp. 499–508, 2000.

[26] F. Vogelsang, M. Kohnen, H. Schneider, F. Weiler, M. W. Kilbinger, B. B. Wein, and R. W. Guenther, "Skeletal maturity determination from hand radiograph by model-based analysis," in *Medical Imaging 2000: Image Processing*, K. M. Hanson, Ed., vol. 3979. SPIE, 2000, pp. 294–305.

[27] M. Kohnen, F. Vogelsang, B. B. Wein, M. W. Kilbinger, R. W. Guenther, F. Weiler, J. Bredno, and J. Dahmen, "Knowledge-based automated feature extraction to categorize secondary digitized radiographs," in *Medical Imaging 2000: Image Processing*, K. M. Hanson, Ed., vol. 3979. SPIE, 2000, pp. 709–717.

[28] M. Kohnen, A. H. Mahnken, A. S. Brandt, S. Steinberg, R. W. Guenther, and B. B. Wein, "Segmentation of the lumbar spine with knowledge-based shape models," in *Medical Imaging 2002: Image Processing*, M. Sonka and J. M. Fitzpatrick, Eds., vol. 4684, no. 1. SPIE, 2002, pp. 1578–1587.

[29] G. Zamora, H. Sari-Sarraf, and L. R. Long, "Hierarchical segmentation of vertebrae from x-ray images," in *Medical Imaging 2003: Image Processing*, M. Sonka and J. M. Fitzpatrick, Eds., vol. 5032. SPIE, 2003, pp. 631–642.

[30] 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*, vol. 6, no. 1, pp. 47–62, 2002.

[31] M. Seise, S. J. McKenna, I. W. Ricketts, and C. A. Wigderowitz, "Double contour active shape models," in *British Machine Vision Conference*, vol. 2, 2005, pp. 159–168.

[32] 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, Tech. Rep. UU-CS-2003-013, 2003.

[33] T. F. Cootes, C. J. Taylor, and A. Lanitis, "Active shape models: Evaluation of a multi-resolution method for improving search," in *British Machine Vision Conference*, 1994, pp. 327–336.

[34] M. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, C. M. Bishop and B. J. Frey, Eds., 2003.

[35] H. H. Thodberg and H. Olafsdottir, "Adding curvature to minimum description length shape models," in *British Machine Vision Conference*, 2003, pp. 251–260.

[36] R. H. Davies, “Learning shape: Optimal models for analysing natural variability,” Ph.D. dissertation, University of Manchester, 2002.
[37] C. E. Rasmussen and J. Quiñonero-Candela, “Healing the relevance vector machine through augmentation,” in *Proceedings of the 22nd International Conference on Machine Learning*, L. De Raedt and S. Wrobel, Eds., 2005, pp. 689–696.