This paper appears in Medical Image Understanding and Analysis, Manchester 2006

# Probabilistic Segmentation of the Knee Joint from X-ray Images

Matthias Seise[1*], Stephen J. McKenna[1], Ian W. Ricketts[1] and Carlos A. Wigderowitz[2]

[1]Division of Applied Computing, University of Dundee, Dundee, Scotland, DD1 4HN
[2]Orthopaedic and Trauma Surgery, Ninewells Hospital and Medical School, Dundee, Scotland, DD1 9SY

**Abstract.** A probabilistic method is proposed for segmentation of the knee joint. A likelihood function is formulated that explicitly models overlapping object appearance. Priors on global appearance and geometry (including shape) are learned from example images. Markov chain Monte Carlo methods are used to obtain samples from a posterior distribution over model parameters from which expectations can be estimated. The result is a probabilistic segmentation that quantifies uncertainty so that measurements such as joint space can be made with associated uncertainty. Joint space area and mean point-to-contour distance are used for evaluation.

## 1 Introduction

The aim of this paper is to outline a probabilistic, model-based segmentation method for the knee joint from x-ray images and to make explicit the uncertainty in the segmentation so obtained. The method explicitly handles the possible overlapping of femur and tibia and their appearance models. Such cases are not handled in methods based on active contours [1], active shape models or active appearance models [2], for example.

Segmentation of objects is often only an intermediate result. Consider for example medical image analysis tasks which involve measuring the size of anatomical structures. Most standard segmentation algorithms result in a single solution without any information as to the confidence in this solution. No information about uncertainty is propagated to the subsequent size estimation step. Furthermore, anatomical structures almost inevitably overlap. In medical applications, especially, it is desirable to have an indication of the certainty of a measurement and to cope with structures that overlap or are in close proximity. The performance of the proposed method is evaluated by applying it to the segmentation of the knee joint to enable the measurement of joint space, an important biomarker for the assessment of osteoarthritis [3].

## 2 Modelling Knee Radiographs

The general task of segmenting modelled objects from an image can be described in a Bayesian framework as that of inferring the conditional distribution $P(\mathcal{M}|\mathcal{I},\mathbb{I},\mathbb{S})$ which is the probability of the model parameters, $\mathcal{M}$, given a test image $\mathcal{I}$, a set of training images, $\mathbb{I}$, and their annotations, $\mathbb{S}$. Each object can be described by its geometry and appearance in the image and, more specifically, in terms of shape parameters, $S$, geometric (non-shape) parameters, $G$, global appearance parameters, $A_g$, and local appearance parameters, $A_l$. Using Bayes' rule and assuming $P(\mathcal{I})$ is fixed:

$$P(\mathcal{M}|\mathcal{I}) \propto p(\mathcal{I}|\mathcal{M})P(\mathcal{M}) = p(\mathcal{I}|G,S,A_g,A_l)P(G)P(S)P(A_g)P(A_l) \tag{1}$$

Assuming pixel-wise conditional independence, the likelihood factors as

$$p(\mathcal{I}|G,S,A_g,A_l) = \prod_{n=1}^{N}\prod_{x\in\Omega_n} p_n(\mathcal{I}(x)|G,S,A_g,A_l) \tag{2}$$

where the likelihood is modelled as not depending on pixels outside a finite image support $\Omega = \bigcup_{n=1}^{N}\Omega_n \subset \mathbb{R}^2$ consisting of $N$ non-intersecting regions.

Point distribution models were used to model shape deformations in femoral and tibial contours [2]. Specifically, each normalised femoral shape $z$ was represented using shape parameters $S_f$ by $z \approx \overline{z} + \Phi(\Lambda^{\frac{1}{2}} \bullet S_f)$ with $\overline{z}$ being the mean shape, $\Phi$ the eigenvectors and $\Lambda^{\frac{1}{2}}$ the square roots of the corresponding eigenvalues of the covariance matrix, all learned from training data. $(\Lambda^{\frac{1}{2}} \bullet S_f)$ denotes the componentwise product. Shape parameters, $S_t$, for tibia were defined analogously. Further geometric parameters were global parameters $G_g = \{t_x, t_y, \theta, s_f\}$ where $(t_x, t_y)$ is the midpoint of the line segment connecting the centres of mass of the femur and tibia, $\theta$ is the rotation of this line segment, and $s_f$ is scale, and local parameters $G_l = \{d, s_t, \theta_f, \theta_t\}$ where $d$ is the separation of the bones, $s_t$ is the scale of the tibia relative to the femur, and $\theta_f$ and $\theta_t$ are local rotations of the femur and tibia (see Figure 1)

X-ray images can vary a lot in terms of brightness and contrast as well as in the imaged region of anatomy and surrounding area. Rather than an appearance model that accounts for the entire image it is sensible to consider only a limited region around the bone contours. Specifically, appearance models were bands of width $2h$ centred on the contour. These can be described by warp functions $\omega : x \in \mathbb{R}^2 \mapsto (\alpha, \delta) \in [0,1] \times [-h, h]$ from subsets of image coordinates to rectangular regions, the texture patches. The value of $\alpha$ corresponds to the normalised arclength of the shape and $\delta$ to the distance normal to the shape. It is important to note that $\omega$ is dependent on the shape and position of the object. Global appearance parameters, brightness $\beta$ and contrast $\gamma$, are included in the model parameters as $A_g = \{\beta, \gamma\}$ and are thus automatically inferred by the segmentation algorithm.

*Corresponding author: M. Seise. email: mseise@computing.dundee.ac.uk

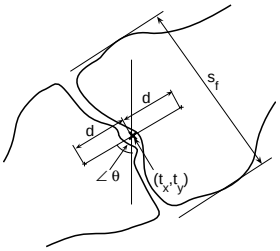



**Figure 1.** Geometric model parameters

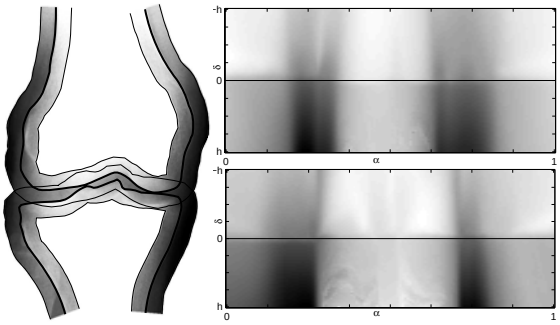



**Figure 2.** Local appearance. Left: masks. Right: associated texture patches (means only)

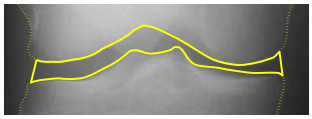


**Figure 3.** Definition of joint space area.

### 2.1 Single Object Appearance

A warp $\omega$ maps intensity values of an image region $\Omega$ onto the rectangular texture patch $T : [0,1] \times [-h,h] \rightarrow \mathbb{R}$ by $T(\omega(\Omega)) = \mathcal{I}(\Omega)$. Making the assumption that pixel values on a texture patch are independently Gaussian distributed, the local appearance is modelled by pixelwise means and variances, $A_l = \{\boldsymbol{\mu}_T, \boldsymbol{\sigma}_T^2\}$ which are estimated from training data using maximum likelihood[1]. The likelihood of the parameters is modelled using a product of Gaussians:

$$p(\mathcal{I}|G,S,A_g,A_l) = \left( \prod_{x \in \Omega} p\left( \frac{\mathcal{I}(x)-\beta}{\gamma} \middle| \boldsymbol{\mu}_T(\omega(x)), \boldsymbol{\sigma}_T^2(\omega(x)) \right) \right)^{\frac{W}{|\Omega|}} \tag{3}$$

where geometry and shape are accounted for by the warp $\omega$ and to avoid the need to always use the same number of observations, the function is normalised with $W$ chosen according to a notional mask size. This weighting enables the importance of the prior to be varied. This was used subsequently in the MCMC algorithm to achieve sensible acceptance rates.

### 2.2 Multiple Object Appearance

When segmenting multiple objects, appearance overlap should be taken into account. Each object's appearance mask divides the image into three image regions. The region corresponding to the mask divides into regions hypothesised as interior and exterior to the object. The third image region is the background not corresponding to the mask. These regions are named $\Omega_f^{in}, \Omega_f^{out}, \Omega_f^{bgd}$ for the femur and $\Omega_t^{in}, \Omega_t^{out}, \Omega_t^{bgd}$ for the tibia. Considering object overlaps, there are nine possible region types: $\Omega_f^{in} \cap \Omega_t^{in}, \Omega_f^{in} \cap \Omega_t^{out}, \ldots, \Omega_f^{bgd} \cap \Omega_t^{bgd}$. For each of these possibilities, likelihood functions must be chosen. Those used here are motivated by x-ray image formation.

#### 2.2.1 X-ray image formation

Photons emitted from an x-ray source and interacting with tissue are absorbed, scattered or transmitted. Noise due to scattering is largely filtered out in modern x-ray radioscopic equipment. The probability $P$ that a photon is absorbed follows Beer's Law, so $P = e^{-m}$ where $m = \int_L \xi(x)dx$, with $\xi$ being the opacity and $L$ the path of the photon through the object. In the simplest case $m = \xi \cdot l$ with $l$ being the thickness of the object and $\xi$ the mean opacity. The intensity $I$ of a point in an x-ray image created by a ray that has passed through an object is then approximated as $I = I_0 e^{-m}$ where $I_0$ is the intensity of a pixel created by an undisturbed ray. (The film, equipment and digitisation process also influence the image but are not modelled here.) Consider the case of two objects that are partially overlapping. An x-ray passing only through the first object creates an intensity $I_1 = I_0 e^{-m_1}$ and one passing only through the second object creates $I_2 = I_0 e^{-m_2}$. The intensity $I_{1,2}$ where the objects' images overlap is then

$$I_{1,2} = I_0 e^{-(m_1+m_2)} = \frac{I_1 I_2}{I_0}$$

Without loss of generality let the maximum intensity in an image $\mathcal{I} : \mathbb{R}^2 \rightarrow [0,1]$ be 1. It can be assumed that at least one pixel in an image is created by an undisturbed ray. Therefore, $I_0 = 1$ and intensities in regions of overlap are approximately multiplicative.

[1]For $N$ texture patches, $\{T_1, \ldots, T_N\}$, $\boldsymbol{\mu}_T$ is defined as the "pixelwise" mean $\boldsymbol{\mu}_T(\alpha,\delta) = \frac{1}{N}\sum_{n=1}^{N} T_n(\alpha,\delta)$ and $\boldsymbol{\sigma}_T^2$ as the "pixelwise" variance $\boldsymbol{\sigma}_T^2(\alpha,\delta) = \frac{1}{N-1}\sum_{n=1}^{N}(T_n(\alpha,\delta) - \boldsymbol{\mu}_T(\alpha,\delta))^2$.

| $x \in$ | Likelihood factor | abbr.[a] |
|---|---|---|
| $\Omega_1$ | $p\left(\frac{\mathcal{I}(x)-\beta}{\gamma}\middle\vert \boldsymbol{\mu}_{T_f}(\omega(x)), \boldsymbol{\sigma}^2_{T_f}(\omega(x))\right)$ | $p_f(x)$ |
| $\Omega_2$ | $p\left(\frac{\mathcal{I}(x)-\beta}{\gamma}\middle\vert \boldsymbol{\mu}_{T_t}(\omega(x)), \boldsymbol{\sigma}^2_{T_t}(\omega(x))\right)$ | $p_t(x)$ |
| $\Omega_3$ | $\frac{1}{2}p\left(\frac{\mathcal{I}(x)-\beta}{\gamma}\middle\vert \boldsymbol{\mu}_{T_f}(\omega(x)), \boldsymbol{\sigma}^2_{T_f}(\omega(x))\right) + \frac{1}{2}p\left(\frac{\mathcal{I}(x)-\beta}{\gamma}\middle\vert \boldsymbol{\mu}_{T_t}(\omega(x)), \boldsymbol{\sigma}^2_{T_t}(\omega(x))\right)$ | $p_{f+t}(x)$ |
| $\Omega_4$ | $\mathcal{F}(\boldsymbol{\mu}_{T_f}(\omega(x)), \boldsymbol{\sigma}^2_{T_f}(\omega(x)), \boldsymbol{\mu}_{T_t}(\omega(x)), \boldsymbol{\sigma}^2_{T_t}(\omega(x)))$ | $p_{\mathcal{F}}(x)$ |

[a] Abbreviation to simplify further notation.

**Table 1.** Likelihood calculation with overlaps ($\mathcal{F}$ is the normal product distribution)

#### 2.2.2 Likelihood functions

The mapping from an observation's region type to its associated likelihood function is given in Table 1. Since the opacity of soft tissue is smaller than that of bone, the likelihood for pixels lying in an overlap of soft tissue and bone is approximated as being the same as if only bone were present. In regions of overlap of the exterior parts of two object masks, a Gaussian mixture (abbreviated as $p_{f+t}(x)$) is used since the observations are modelled by both masks. In regions of overlap of the interior parts of two object masks, the intensities are approximately multiplicative, so the *normal product distribution* $\mathcal{F}$ is used. Usually its density *cannot* be calculated analytically but can be approximated by $\mathcal{N}(\boldsymbol{\mu}_{T_f}(\omega(x))\boldsymbol{\sigma}^2_{T_f}(\omega(x)), \boldsymbol{\sigma}^2_{T_f}(\omega(x))\boldsymbol{\sigma}^2_{T_t}(\omega(x)) + \boldsymbol{\mu}_{T_f}(\omega(x))\boldsymbol{\sigma}^2_{T_t}(\omega(x)) + \boldsymbol{\sigma}^2_{T_f}(\omega(x))\boldsymbol{\sigma}^2_{T_t}(\omega(x)))$ [4]. The final likelihood is then

$$p(\mathcal{I}|G, S, A_g, A_l) = \left(\prod_{x\in\Omega_1} p_f(x) \cdot \prod_{x\in\Omega_2} p_t(x) \cdot \prod_{x\in\Omega_3} p_{f+t}(x) \cdot \prod_{x\in\Omega_4} p_{\mathcal{F}}(x)\right)^{\frac{W}{V}} \tag{4}$$

with

$$\Omega_1 = (\Omega_f^{in} \cap (\Omega_t^{out} \cup \Omega_t^{bgd})) \cup (\Omega_f^{out} \cap \Omega_t^{bgd}) \qquad \Omega_2 = (\Omega_t^{in} \cap (\Omega_f^{out} \cup \Omega_f^{bgd})) \cup (\Omega_t^{out} \cap \Omega_f^{bgd})$$

$$\Omega_3 = \Omega_f^{out} \cap \Omega_t^{out} \qquad \Omega_4 = \Omega_f^{in} \cap \Omega_t^{in} \qquad V = \left|\bigcup_{i=1}^{4} \Omega_i\right|$$

**Learning the likelihood model** Appearance learning constructs the texture patches $\{T_1, \ldots, T_n\}$ and calculates the pixelwise means $\{\boldsymbol{\mu}_T\}$ and variances $\{\boldsymbol{\sigma}^2_T\}$. For our application, only regions not overlapping other bones were used to estimate them. Missing data were then estimated by interpolating from neighbouring pixels. Estimated means are shown in Figure 2.

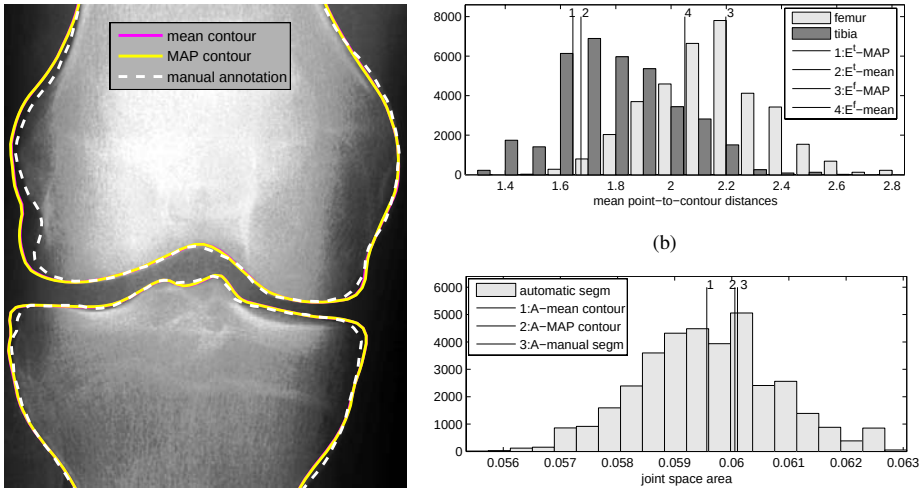
### 2.3 Prior Distributions

Priors $P(G), P(S), P(A_g), P(A_l)$ were chosen to reflect beliefs about the behaviour of the variables. Gaussian priors were used for brightness $\beta$, contrast $\gamma$, rotations $\theta, \theta_f, \theta_t$ and local appearance $A_l$ with the corresponding means and variances estimated from training data using maximum likelihood. Translation $(t_x, t_y)$ had a uniform prior over a rectangular region of the image and scale priors for $s_f$ and $s_t$ were log-Gaussian. The specifying parameters of the latter distributions were also estimated from training data.

The learning of the shape model priors was similar to the ASM shape model building step [2] and was performed independently for femur and tibia. Femur shapes $\mathbb{S}_f = \{s_f^1, \ldots, s_f^N\}$ were aligned and normalised using Procrustes analysis. The mean $\overline{S}_f$ and the covariance matrix $\Sigma_f$ were calculated. PCA was applied to $\Sigma_f$ decomposing it into $\Sigma_f = \Lambda_f \cdot \Phi_f \cdot \Lambda_f^{-1}$, the eigenvalue vector $\Lambda_f$ and the eigenvectors $\Phi_f$. With the standard assumption of a Gaussian shape space it is implied that the prior of the shape parameters $S_f$ is the multivariate standard normal distribution. Tibia shapes $\mathbb{S}_t$ were processed analogously.

### 2.4 Sampling the Posterior with Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are used to generate samples from a target distribution and subsequently to estimate expectations of functions under this distribution (see Andrieu *et al.* [5] for a good introduction). To sample from the posterior, Equation (1), we used a mixture of Metropolis-Hastings random walk MCMC and a variant of hybrid MCMC. Geometric parameters, $G$, were updated componentwise and shape parameters blockwise using Gaussian random walk. Blocks were chosen such that each block accounted for approximately the same variance. Global appearance parameters $A_g = \{\beta, \gamma\}$ were updated together using the Langevin method since partial derivatives with respect to brightness $\frac{\partial P(\mathcal{I}|\mathcal{M})}{\partial \beta}$ and contrast $\frac{\partial P(\mathcal{I}|\mathcal{M})}{\partial \gamma}$ are easily and efficiently calculated.




(a) (b) (c)

**Figure 4.** (a) Example segmentation result with corresponding histograms of (b) mean point-to-contour distances and (c) joint space areas for $40,000$ MCMC samples.

## 3 Experiments

Methods were evaluated on a data set of 30 standard clinical x-rays (see also [3]). 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. Shapes were brought into correspondence using the MDL approach [6] with curvature [7]. After applying PCA, the 16 major modes were left unchanged. Rather than discarding the minor modes, their variances were set to the mean value of the remaining variances to enable search outside the lower-dimensional subspace.

A coarse-to-fine algorithm was used for speed since likelihood calculation based on this model was computationally expensive. A two-step approach was used to first find the main modes of the posterior at low resolution and then explore them at full resolution.

**1st step: coarse resolution.** Shape was fixed to the mean shape while the other geometric and the global appearance parameters were updated. The Markov chain was initialised randomly as a sample from the prior distribution. The standard deviations of the proposal functions were set equal to the standard deviations of the associated parameters learned from the training set. The weight $W$ was set so that likelihood and prior were of comparable magnitude so as to encourage exploration of the posterior. (Typical performance: 20% acceptance rate, 8 minutes per 1000 samples on Centrino 1.9 GHz, 1GB RAM, implementation in Matlab.)

**2nd step: fine resolution.** The chain was initialised at the MAP solution obtained from step 1. The weight $W$ was increased to encourage exploration of local modes. All model parameters were updated, geometric and global appearance parameters with decreased variance. Shape parameters were updated in blocks with proposal standard deviations set to half the standard deviations of the respective parameters. Further improvements in speed were achieved by sampling pixels in "uninteresting" regions of the mask at a lower density. The application of measuring the joint space demands higher accuracy along the femoral condyles and tibial plateaux than elsewhere. Therefore, the likelihood was calculated for every pixel around the joint space and only at quarter resolution in the other areas. Step 2 resulted in a representation of the posterior distribution (with the first $10\%$ of samples discarded as burn-in). (Typical performance: $8\%$ acceptance rate, 30 minutes per 1000 samples on Centrino 1.9 GHz, 1GB RAM, implementation in Matlab.)

Figure 4(a) shows an example test image with the mean and MAP estimates overlaid along with a manual annotation. The mean and MAP contours cannot be readily distinguished and this was also true of most other test images. Inferred contours can be compared to a manually annotated contour based on *mean point-to-contour distance* defined as the average Euclidean distance from the landmark positions to the annotated contour. Mean point-to-contour distances along the femoral condyles ($E^f$) and along the tibial plateaux ($E^t$) of the image in Figure 4(a) are shown in Figure 4(b). The values of $E^f$ and $E^t$ are also indicated for the mean and MAP contours computed from the Monte Carlo samples. Further evaluation is based on the joint space area. In the absence of a generally accepted definition of joint space area we used the area between the tibial plateaux and femoral condyles as shown in Figure 3, normalised by the overall scale of the bone. Figure 4(c) shows the calculated areas as well as those of the mean contour, the MAP contour and the manual annotation for the image in Figure 4(a).

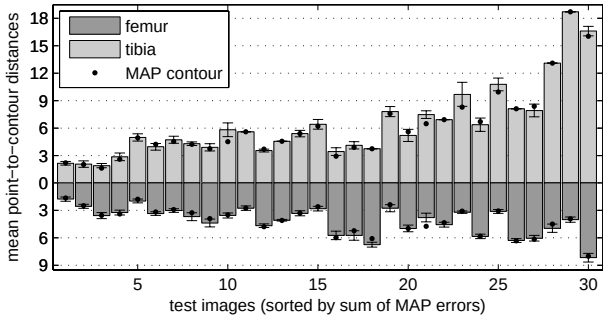



**Figure 5.** Statistics summarising mean point-to-contour distances for the test images (see text).

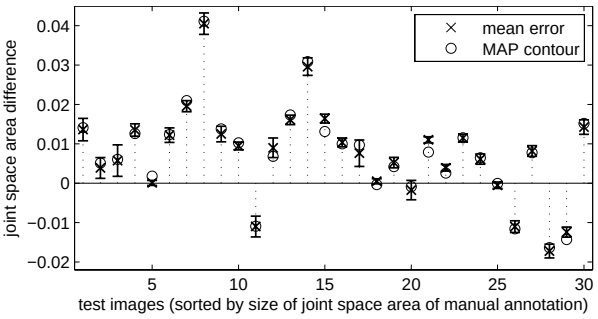



**Figure 6.** Statistics summarising difference in joint space area measurements (see text).

The values of $E^f$ and $E^t$ obtained with the thirty test images are summarised in Figure 5. Results for the femur are shown below the baseline and for the tibia above. Each histogram bar indicates the Monte Carlo estimate of the expectation of the mean point-to-contour distance under the posterior distribution. Error bars indicate standard deviations and MAP contour errors are also indicated. Analogously, the differences between the the joint space area of the manual annotation and of the Monte Carlo samples and the MAP contour are summarised in Figure 6. It shows that the automatic method tends to give larger areas than the human annotator.

## 4 Discussion and Conclusions

A probabilistic framework for segmentation of the knee joint was described in which appearance overlaps were explicitly modelled. The output was not a single contour but Monte Carlo samples from a posterior distribution over contours. The approach is broadly applicable since overlapping objects are almost inevitable in medical applications (e.g. [8]). Furthermore, since segmentation is often only an intermediate step it is important to quantify and propagate uncertainty to higher level processes. An important example is the estimation of areas and volumes in medical images. The method was applied to segmentation of femoral condyles and tibial plateaux and the joint space area was computed as a possible proxy for cartilage volume. It achieved reasonable results based on a relatively small and challenging data set. The accuracy of the joint space area measurement has to be confirmed in further studies. The current implementation would seem to have a tendency to overestimate joint space. However, it must be noted that, so far, neither errors in the manual annotation nor flaws in the definition of the joint space area were considered. Future work could extend the approach by explicitly modelling uncertainty in the manual annotations upon which learning is performed. Other opportunities include the use of data-driven MCMC schemes to ensure full exploration of the multi-modal posterior [9] and enhancement of the likelihood model.

## 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. M. Kass, A. Witkin & D. Terzopoulos. “Snakes: Active contour models.” *IJCV* **1(4)**, pp. 321–331, 1988.
2. T. F. Cootes & C. J. Taylor. “Statistical models of appearance for computer vision.” Tech. rep., University of Manchester, 2004.
3. M. Seise, S. J. McKenna, I. W. Ricketts et al. “Double contour active shape models.” In *BMVC*, pp. 159 – 168. 2005.
4. R. Ware & F. Lad. “Approximating the distribution for sums of products of normal variables.” Tech. rep., University of Canterbury, 2003.
5. C. Andrieu, N. de Freitas, A. Doucet et al. “An introduction to MCMC for machine learning.” *Machine Learning* **50**, pp. 5–43, Jan-Feb 2003.
6. R. H. Davies, C. Twining, T. F. Cootes et al. “A minimum description length approach to statistical shape modelling.” *IEEE Trans. Med. Imag.* **21(5)**, pp. 525 – 537, 2002.
7. H. H. Thodberg & H. Olafsdottir. “Adding curvature to minimum description length shape models.” In *BMVC*, pp. 251 – 260. 2003.
8. P. P. Smyth, C. J. Taylor & J. E. Adams. “Vertebral shape: automatic measurement with active shape models.” *Radiology* **211(2)**, pp. 571 – 578, 1999.
9. C. Liu, H.-Y. Shum & C. Zhang. “Hierarchical shape modeling for automatic face localization.” In *ECCV*, pp. 687–703. Springer-Verlag, London, UK, 2002.