

Segmenting Multiple Objects with Overlapping Appearance and Uncertainty

Matthias Seise¹, Stephen J. McKenna¹, Ian W. Ricketts¹
and Carlos A. Wigderowitz²

¹Division of Applied Computing, University of Dundee, UK

²Orthopaedic & Trauma Surgery, Ninewells Hospital, Dundee, UK

{mseise, stephen, ricketts}@computing.dundee.ac.uk, cawigderowitz@lineone.net
www.computing.dundee.ac.uk/projects/Vision

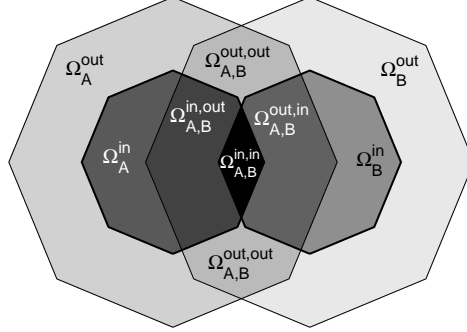
Abstract

A probabilistic method is proposed for segmentation of multiple objects that overlap or are in close proximity to one another. 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 method is described in detail for the problem of segmenting femur and tibia in x-ray images. The result is a probabilistic segmentation that quantifies uncertainty so that measurements such as joint space can be made with associated uncertainty.

1 Introduction

The aim of this paper is to outline a probabilistic model-based segmentation method for multiple, possibly overlapping, objects and to make explicit the uncertainty in the segmentation so obtained. The method explicitly handles objects that overlap, or whose appearance models overlap, in the image. Such cases are *not* handled in methods based on active contours [6], active shape models or active appearance models [3], for example. The method is applied here to analysis of radiographic images where overlapping objects result in pixel values that depend on multiple objects. A somewhat related approach for segmenting multiple *opaque* objects was reported by Li [7] who posed the object matching as a labeling problem which was solved by a MAP estimation.

Segmentation of objects is often only an intermediate result. Consider for example medical image analysis tasks involving 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 proposed method is evaluated for the segmentation of the knee joint for the purpose of the measurement of joint space which is an important biomarker for the assessment of osteoarthritis [10].

Figure 1: Regions in overlapping objects A and B

2 Probabilistic Segmentation

The general task of segmenting modelled objects from an image can be described in the Bayesian framework as that of inferring the conditional distribution $P(\mathcal{M}|\mathcal{J}, \mathbb{I}, \mathbb{S})$ which is the probability of the model parameters, \mathcal{M} , given a test image \mathcal{J} , 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 $P(\mathcal{J})$ is fixed:

$$\begin{aligned} P(\mathcal{M}|\mathcal{J}) &\propto p(\mathcal{J}|\mathcal{M})P(\mathcal{M}) \\ &= p(\mathcal{J}|G, S, A_g, A_l)P(G)P(S)P(A_g)P(A_l) \end{aligned} \quad (1)$$

Assuming pixel-wise conditional independence, the likelihood factors as

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

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. In general, different likelihood functions will be used at different locations with respect to the object. Furthermore, likelihood functions for a region in which object appearances overlap can be modelled by combining object-specific likelihood factors appropriately. An object divides its image into distinct regions. Overlapping objects result in further distinct regions. Figure 1 illustrates an example in which two convex objects, A and B , overlap. Their appearance models have finite support, in this case an interior region and an exterior region within a fixed distance of the object boundary. Eight regions arise, each potentially requiring different forms of likelihood functions. These regions are $\{\Omega_A^{in}, \Omega_B^{in}, \Omega_A^{out}, \Omega_B^{out}, \Omega_{A,B}^{in,in}, \Omega_{A,B}^{in,out}, \Omega_{A,B}^{out,in}, \Omega_{A,B}^{out,out}\}$, denoting inside A , inside B , outside A , outside B , inside both, inside A but outside B , outside A but inside B , and outside both, respectively. Note that regions are not always contiguous. The manner in which their likelihood functions are computed depends on the actual application.

3 Modelling Knee Radiographs

The method is now specified fully for the case of segmentation of multiple bones from plane radiographs. We focus specifically on femur and tibia, segmentations of which are

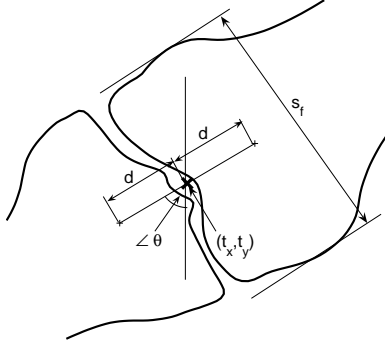
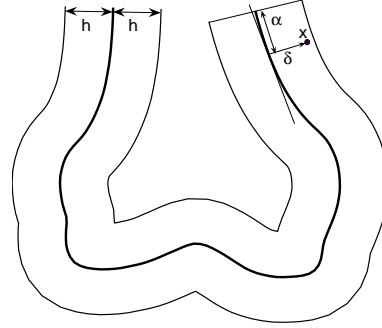


Figure 2: Geometric model parameters

Figure 3: Mask and warp $\omega : x \mapsto (\alpha, \delta)$

important in the assessment of osteoarthritis via measurement of joint space.

Point distribution models were used to model shape deformations in each bone's contours [3]. Specifically, each normalised femoral shape x was represented using shape parameters S_f by $x \approx \bar{x} + \Phi \Lambda^{\frac{1}{2}} S_f$ with \bar{x} being the mean shape, Φ the eigenvectors and $\Lambda^{\frac{1}{2}}$ the square roots of the corresponding eigenvalues of the covariance matrix, all learned from training data. 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, θ 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 θ_f and θ_t are local rotations of the femur and tibia (see Figure 2).

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 therefore makes sense to consider only a limited region around the bone contours. Specifically, appearance models were bands of width $2h$ centred on the contour. Mathematically, 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. As shown in Figure 3, the value of α corresponds to the normalised arclength of the shape and δ to the distance normal to the shape. It is important to note that ω is dependent on the shape and position of the object. Global appearance parameters, brightness β and contrast γ , are included in the model parameters as $A_g = \{\beta, \gamma\}$ and are thus automatically inferred by the segmentation algorithm.

3.1 Single Object Appearance

A warp ω maps intensity values of an image region Ω onto a rectangular texture patch

$$T : [0, 1] \times [-h, h] \rightarrow \mathbb{R} \text{ by } T(\omega(\Omega)) = \mathcal{J}(\Omega).$$

Making the assumption that pixel values on a texture patch are independently Gaussian distributed, the local appearance is modelled by means and variances, $A_l = \{\mu_T, \sigma_T^2\}$ which are estimated from training data using maximum likelihood. The likelihood of the parameters given the image \mathcal{J} is modelled as:

$$p(\mathcal{J}|G, S, A_g, A_l) = \prod_{x \in \Omega} p\left(\frac{\mathcal{J}(x) - \beta}{\gamma} \middle| \mu_T(\omega(x)), \sigma_T^2(\omega(x))\right) \quad (3)$$

where geometry and shape are accounted for by the warp ω .

3.2 Multiple Object Appearance

When segmenting multiple objects, appearance overlap should be taken into account. Here the two-object case is considered but generalisation to more objects is straightforward. 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: $\Omega_f^{in} \cap \Omega_t^{in}, \Omega_f^{in} \cap \Omega_t^{out}, \dots, \Omega_f^{bgd} \cap \Omega_t^{bgd}$. For each of these regions, the nature of the likelihood functions must be considered. Those used here were motivated by x-ray image formation.

3.2.1 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 $m = \int_L \mu(x) dx$, with μ being the opacity and L the path of the photon through the object. In the simplest case $m = \mu \cdot l$ with l being the thickness of the object and μ 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

$$I = I_0 e^{-m} \quad (4)$$

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

$$I_{1,2} = I_0 e^{-(m_1+m_2)} = \frac{I_1 I_2}{I_0} \quad (5)$$

Without loss of generality let the maximum intensity in an image $\mathcal{J} : \mathbb{R}^2 \rightarrow [0, 1]$ be 1. Note that x-ray images are negatives, so an intensity of 1 is visualised as black. It can be assumed that at least one pixel in an image is created by an undisturbed ray. Therefore, intensities in regions of overlap are approximately multiplicative.

3.2.2 Likelihood functions

The mapping from an observation's region type to its associated form of likelihood function is given in Table 1. Since the opacity of soft tissue is less than that of bone, the likelihood for pixels lying in the region of 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* $\mathcal{F}(\mu_{T_f}(\omega(x)), \sigma_{T_f}^2(\omega(x)), \mu_{T_t}(\omega(x)), \sigma_{T_t}^2(\omega(x)))$ is used. The normal product distribution is defined as follows:

Let X be normally distributed with $\mathcal{N}(\mu_X, \sigma_X^2)$ and Y with $\mathcal{N}(\mu_Y, \sigma_Y^2)$. XY is then *normal product* distributed, denoted by $XY \sim \mathcal{F}(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2)$.

$x \in$	Likelihood factor	abbr. ^a
$\Omega_f^{in} \cap \Omega_t^{out}$	$p \left(\frac{\mathcal{J}(x) - \beta}{\gamma} \middle \mu_{T_f}(\omega(x)), \sigma_{T_f}^2(\omega(x)) \right)$	$p_f(x)$
$\Omega_f^{in} \cap \Omega_t^{bgd}$	$p \left(\frac{\mathcal{J}(x) - \beta}{\gamma} \middle \mu_{T_f}(\omega(x)), \sigma_{T_f}^2(\omega(x)) \right)$	$p_f(x)$
$\Omega_f^{out} \cap \Omega_t^{bgd}$	$p \left(\frac{\mathcal{J}(x) - \beta}{\gamma} \middle \mu_{T_f}(\omega(x)), \sigma_{T_f}^2(\omega(x)) \right)$	$p_f(x)$
$\Omega_f^{out} \cap \Omega_t^{in}$	$p \left(\frac{\mathcal{J}(x) - \beta}{\gamma} \middle \mu_{T_t}(\omega(x)), \sigma_{T_t}^2(\omega(x)) \right)$	$p_t(x)$
$\Omega_f^{bgd} \cap \Omega_t^{in}$	$p \left(\frac{\mathcal{J}(x) - \beta}{\gamma} \middle \mu_{T_t}(\omega(x)), \sigma_{T_t}^2(\omega(x)) \right)$	$p_t(x)$
$\Omega_f^{bgd} \cap \Omega_t^{out}$	$p \left(\frac{\mathcal{J}(x) - \beta}{\gamma} \middle \mu_{T_t}(\omega(x)), \sigma_{T_t}^2(\omega(x)) \right)$	$p_t(x)$
$\Omega_f^{in} \cap \Omega_t^{in}$	$\mathcal{F}(\mu_{T_f}(\omega(x)), \sigma_{T_f}^2(\omega(x)), \mu_{T_t}(\omega(x)), \sigma_{T_t}^2(\omega(x)))$	$p_{\mathcal{F}}(x)$
$\Omega_f^{out} \cap \Omega_t^{out}$	$\frac{1}{2} p \left(\frac{\mathcal{J}(x) - \beta}{\gamma} \middle \mu_{T_f}(\omega(x)), \sigma_{T_f}^2(\omega(x)) \right)$ $+ \frac{1}{2} p \left(\frac{\mathcal{J}(x) - \beta}{\gamma} \middle \mu_{T_t}(\omega(x)), \sigma_{T_t}^2(\omega(x)) \right)$	$p_{f+t}(x)$

^aAbbreviation to simplify further notation.

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

Usually the density \mathcal{F} *cannot* be calculated analytically and approximation techniques have to be used to estimate it [13]. Practically, the following approximation can be used.

$$\mathcal{F}(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2) \approx \mathcal{N}(\mu_X \mu_Y, \sigma_X^2 \sigma_Y^2 + \mu_X \sigma_Y^2 + \mu_Y \sigma_X^2)$$

The final likelihood is then

$$p(\mathcal{J}|G, S, A_g, A_l) = \prod_{x \in \Omega_1} p_f(x) \cdot \prod_{x \in \Omega_2} p_t(x) \cdot \prod_{x \in \Omega_3} p_{f+t}(x) \cdot \prod_{x \in \Omega_4} p_{\mathcal{F}}(x) \quad (6)$$

with

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

$$\Omega_3 = \Omega_f^{out} \cap \Omega_t^{out} \quad \Omega_4 = \Omega_f^{in} \cap \Omega_t^{in}$$

Learning the likelihood model Appearance learning constructs the texture patches $\{T_1, \dots, T_n\}$ and calculates the pixelwise means $\{\mu_T\}$ and variances $\{\sigma_T^2\}$. 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 4.

3.3 Prior Distributions

Priors $P(G), P(S), P(A_g), P(A_l)$ were chosen to reflect beliefs about the behaviour of the variables. In this application, Gaussian priors were used for brightness β , contrast γ , 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 the geometric mean and geometric variance and they were also estimated from training data.

The learning of the shape model priors was performed independently for femur and tibia. Femur shapes $\mathbb{S}_f = \{s_f^1, \dots, s_f^N\}$ were aligned and normalised using Procrustes analysis, the mean \bar{S}_f and the covariance matrix Σ_f calculated. PCA was applied to Σ_f decomposing it into $\Sigma_f = \Lambda_f \cdot \Phi_f \cdot \Lambda_f^{-1}$, the eigenvalue vector Λ_f and the eigenvectors Φ_f . The priors on the shape parameters were modelled as standard Gaussian distributions.

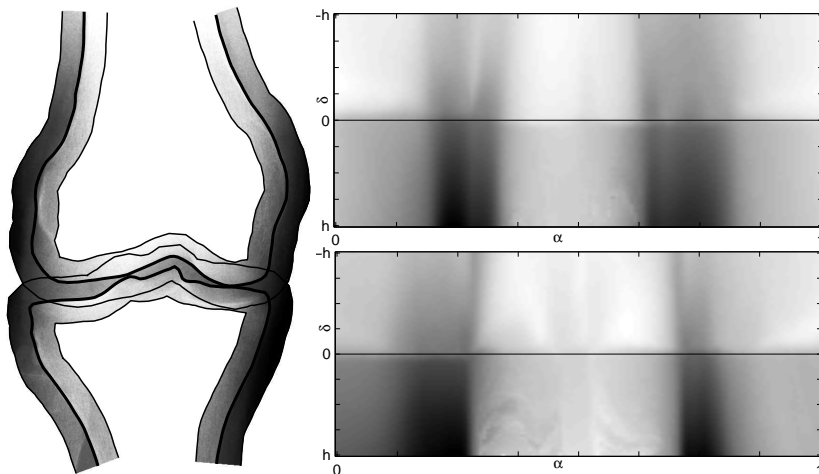


Figure 4: Local appearance. Left: masks. Right: associated texture patches (means only)

3.4 Sampling the Posterior with Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are used to generate samples from a target distribution $\pi(x)$ and subsequently to estimate expectations of functions under this distribution (see Andrieu *et al.* [1] for a good introduction). The target distribution here was the posterior of Equation (1). The Metropolis-Hastings MCMC algorithm iteratively samples from a proposal distribution $q(x'|x)$ and accepts proposals with probability $\min\left\{1, \frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)}\right\}$. Note that it is sufficient to know π up to a normalising constant. The design of the proposal distribution q is important and many practical MCMC algorithms differ only in the choice of q . The simplest choice is a symmetric random walk proposal $q(x'|x) = q(x|x')$, usually with $q \sim \mathcal{N}(x, \Sigma)$. The choice of the covariance matrix Σ is then important since large steps can lead to high rejection rates while small steps result in slow exploration and failure to explore certain modes of the distribution. Convergence can be improved by including gradient information in the proposal. The Langevin method uses the proposal distribution $q(x'|x) = \mathcal{N}(x + 0.5r^2\Delta(x), r^2)$ with $\Delta(x)$ being the gradient and r the expected step size. It is a simple variant of hybrid (Hamiltonian) MCMC. In many computer vision applications it is not possible to compute gradients analytically and this is also true of most parameters in this application. However, partial derivatives with respect to brightness $\frac{\partial P(j|M)}{\partial \beta}$ and contrast $\frac{\partial P(j|M)}{\partial \gamma}$ are easily and efficiently calculable. A useful property of MCMC is that it is possible to combine several samplers into mixtures or cycles. This means, for example, that it is possible to update the parameter vector x componentwise or blockwise, i.e. the proposal distribution affects either one or more components of x . Another possibility is to have global proposals to explore the global behaviour and local proposals to explore local modes. However, one critical property of practical MCMC algorithms is *when* the chain is converging, i.e. how many samples have to be discarded as burn-in. Various diagnostic tools have been developed to assess the convergence, see [2] for a review. Here a rule-of-thumb of discarding 10% as initial burn-in was adopted.

We used a mixture of the above proposals. 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

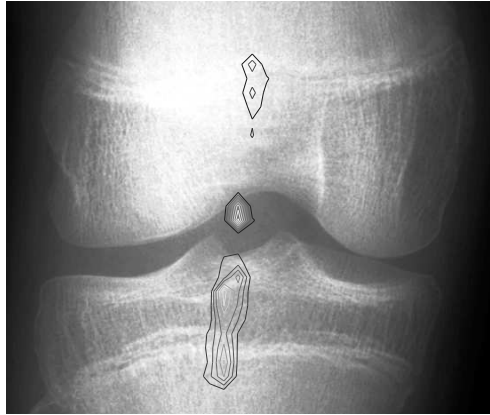


Figure 5: Contour plot of posterior distribution projected onto t_x, t_y .

appearance parameters $A_g = \{\beta, \gamma\}$ were updated together using the Langevin method. The likelihood and also the posterior were quite highly peaked with multiple local maxima and using standard MCMC would have led to very low acceptance rates and thus very slow mixing. To achieve sensible acceptance rates and to make the inference feasible, a variant of tempering (e.g. [5]) to smooth the likelihood and as such the posterior was used.

4 Experiments

The method was evaluated on a data set of 30 standard clinical x-rays (see also [10]). 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 [4] with curvature [12]. 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 global appearance parameters were updated. The chain was initialised randomly, sampled from the prior distributions. The standard deviations of the proposal functions were set equal to the standard deviations of the associated parameters learned from the training set. The log-likelihood was multiplied by a constant so that likelihood and prior were of comparable magnitude so as to encourage exploration of the multi-modal tempered distribution.

2nd step: fine resolution. The chain was initialised at the MAP solution obtained from the first step. 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. The blocks were:

$$\text{eigenmodes } \underbrace{\underbrace{1}_{1^{st}}, \underbrace{2}_{2^{nd}}, \underbrace{3, 4}_{3^{rd}}, \underbrace{5, 6, \dots, 16}_{4^{th}}}_{\text{major eigenmodes}}, \underbrace{17, \dots}_{5^{th} \text{ block}}_{\text{minor eigenmodes}}$$

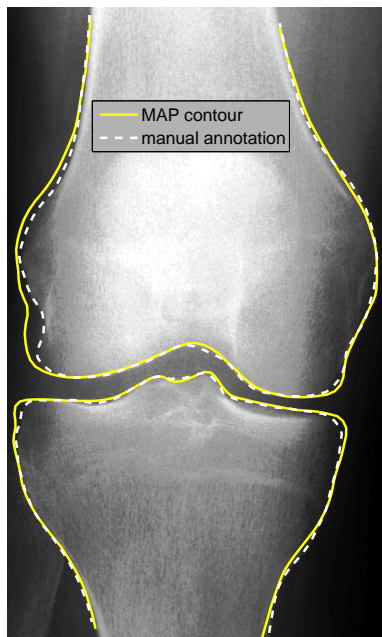


Figure 6: Example segmentation result.

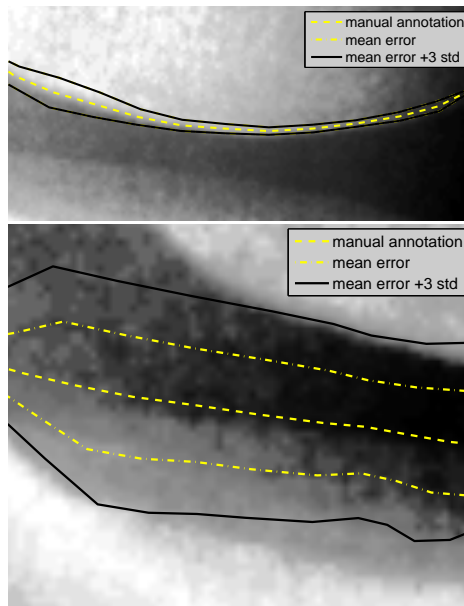


Figure 7: Illustration of uncertainty on segments of the femur (top) and tibia (bottom).

The choice to have four blocks of major eigenmodes and one of the minor eigenmodes was a trade-off between speed and convergence. Large blocks can lead to a higher rejection rate (fewer likelihood evaluations) and small blocks to better convergence since in each iteration only a few shape parameters are changed, but with the cost of many likelihood evaluations. 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.

Figure 5 shows a contour plot of a posterior projected onto the translation parameter space. The posterior appears to be highly peaked in the x -direction while there are ‘false’ modes in the y -direction. However, the second step in the MCMC scheme is likely to explore the central, more highly peaked, mode well.

Figure 6 shows an example test image with the MAP estimates overlaid along with a manual annotation. The mean contour cannot be readily distinguished from the MAP contour and is therefore not shown. 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 6 are shown in Figure 8. The values of E^f and E^t are also indicated for the mean and MAP contours computed from the Monte Carlo samples.

The values of E^f and E^t obtained with the thirty test images are summarised in Figure 9. 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 stan-

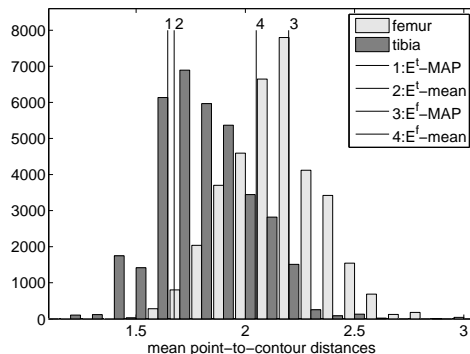


Figure 8: Mean point-to-contour distances for contours obtained from the image in Figure 6.

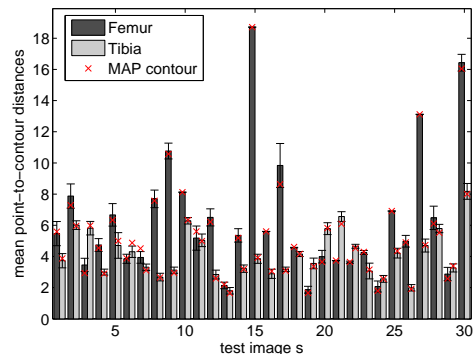


Figure 9: Statistics summarising mean point-to-contour distances for the test images (see text).

standard deviations and MAP contour errors are also indicated.

It is interesting to investigate which parts of the bone contours give rise to large variations in contour localisation. Low variation should indicate good discrimination between the interior and exterior of the object. This reasoning is approximate since points on the contour are updated in blocks depending on the (global) shape model. Figure 7 shows two examples in which the dashed line indicates a manual annotation. The dot-dashed line indicates a *mean distance contour* created by perpendicular displacement of each point on the contour by that point's expected point-to-contour distance under the posterior distribution. The solid line indicates displacement by an additional three standard deviations. The top example shows a part of the femur where point-to-contour distances are small, indicating agreement; a strong edge is apparent. The lower example is a part of the tibia that results in far greater uncertainty.

5 Discussion and Conclusions

A probabilistic framework for segmentation of multiple objects was described in which object appearance overlaps were explicitly modelled. The output is not a single contour but Monte Carlo samples from a posterior distribution over contours. The approach is broadly applicable since, for example, segmentation of overlapping objects is almost inevitable in medical applications (e.g. [11]). Depending on the actual application, different likelihood functions might have to be adopted for example to account for differing imaging modalities. The likelihood functions used here were explicitly defined for x-ray images.

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 as such could be used to estimate joint space as a proxy for cartilage volume, for example. It achieved reasonable results based on a relatively small and challenging data set.

Bayesian inference provides the machinery for quantifying segmentation uncertainty and the method presented here is a step towards that goal. The results showed that the exploration of the multi-modal posterior was suboptimal using the standard MCMC techniques. The use of data-driven and mode-hopping MCMC schemes for better exploration

of the posterior should be investigated [8, 9]. Better models of the shape-free appearance also need to be incorporated in this framework. A significant source of uncertainty is the process of manual annotation, resulting in inter-observer and intra-observer variability. Future work could usefully extend the approach by explicitly modelling uncertainty in the manual annotations upon which learning is performed.

Acknowledgments M. Seise was partially funded by EPSRC. H. H. Thodberg made available the code for the MDL method and Dr B. Oliver provided the knee images. Thanks to Tim Roberts for helpful discussions.

References

- [1] C. Andrieu, N. de Freitas, A. Doucet, and M. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50:5 – 43, 2003.
- [2] S. P. Brooks and G. O. Roberts. Assessing convergence of Markov Chain Monte Carlo algorithms. *Statistics and Computing*, 8(4):319 – 335, 1998.
- [3] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, 2004.
- [4] R. H. Davies, C. Twining, T. F. Cootes, and C. J. Taylor. A minimum description length approach to statistical shape modelling. *IEEE Trans. Med. Imag.*, 21(5):525 – 537, 2002.
- [5] C. Geyer and E. A. Thompson. Annealing Markov Chain Monte Carlo with application to ancestral inference. *J. Amer. Statist. Assoc.*, 90:909 – 920, 1995.
- [6] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321 – 331, 1988.
- [7] S. Z. Li. Recognizing multiple overlapping objects in image: an optimal formulation. *IEEE Trans. Image Processing*, 9(2):273 – 277, 2000.
- [8] C. Liu, H.-Y. Shum, and C. Zhang. Hierarchical shape modeling for automatic face localization. In *ECCV*, pages 687 – 703, London, UK, 2002.
- [9] R. M. Neal. Sampling from multimodal distributions using tempered transitions. *Statistics and Computing*, 6(4):353 – 366, 1996.
- [10] M. Seise, S. J. McKenna, I. W. Ricketts, and C. A. Wigderowitz. Double contour active shape models. In *BMVC*, volume 2, pages 159 – 168, 2005.
- [11] P. P. Smyth, C. J. Taylor, and J. E. Adams. Vertebral shape: automatic measurement with active shape models. *Radiology*, 211(2):571 – 578, 1999.
- [12] H. H. Thodberg and H. Olafsdottir. Adding curvature to minimum description length shape models. In *BMVC*, pages 251 – 260, 2003.
- [13] R. Ware and F. Lad. Approximating the distribution for sums of products of normal variables. Technical report, University of Canterbury, 2003.