Ultrasound Image Segmentation using Stochastic Templates

C. A. Glasbey

Biomathematics and Statistics Scotland, JCMB, King's Buildings, Edinburgh EH9 3JZ, Scotland

Point distribution models (PDMs) are incorporated into Bayesian image analysis, thus combining two approaches to the fitting of stochastic templates. Manually-segmented images are used to identify both a PDM and a likelihood function, leading to a posterior distribution from which inferences can be drawn. The methodology is explored and illustrated using 144 ultrasound images of sheep. A pseudo-likelihood is found to give better results than a likelihood based on the distribution of pixel values in the training images. Estimates of sheep fat and muscle depths are shown to be comparable in accuracy with manual interpretation of images.

Keywords: Bayesian methods, Likelihood, Point distribution models.

Introduction

Templates are examples of objects to be located in images. The use of rigid templates is well established, but is usually inadequate in biological applications where natural variability ensures that shapes are not identical. Stochastic, or deformable, templates generalise rigid templates by allowing variation within prescribed probability distributions. Two methodologies have recently emerged in the machine vision and statistical literature, each with distinct strengths and weaknesses.

- The machine vision approach is based on point distribution models (PDMs), which are extracted from manually-interpreted training images, and used to specify active shape models (ASMs) which are fitted to new images to match edges (Cootes *et al.*, 1995).
- The statistical approach is a form of Bayesian image analysis, where a prior model specifies what is expected in an image. This is combined with the likelihood of an observed image conditional on the model, resulting in a posterior distribution from which inferences can be drawn (Grenander and Miller, 1994, and Phillips and Smith, 1994).

The strength of the machine vision approach is the ease of template formation, whereas the statistical approach provides a powerful set of tools for fitting a template to data.

This paper reports an experiment to gain the benefits of both approaches. In §2, ultrasound images of sheep are introduced, which present a challenging problem for automatic segmentation. In §3, training data, in the form of manually-specified templates, are used to identify a PDM for use as a prior distribution. Bayes' theorem is used to combine the PDM with a likelihood function of simple form, but the resulting posterior distribution has a mode which does not consistently coincide with the parameter values in the original template. Therefore, in §4, a new method is proposed in which the training data are further used to modify the likelihood function. Results for the training data, and for a validation data set, show that sheep fat and muscle depths are estimated as accurately as for manual interpretation. Finally, in §5, generalisations of the method are discussed.

Ultrasound Images

Non-invasive imaging techniques such as ultrasound, X-ray computed tomography (CT) and

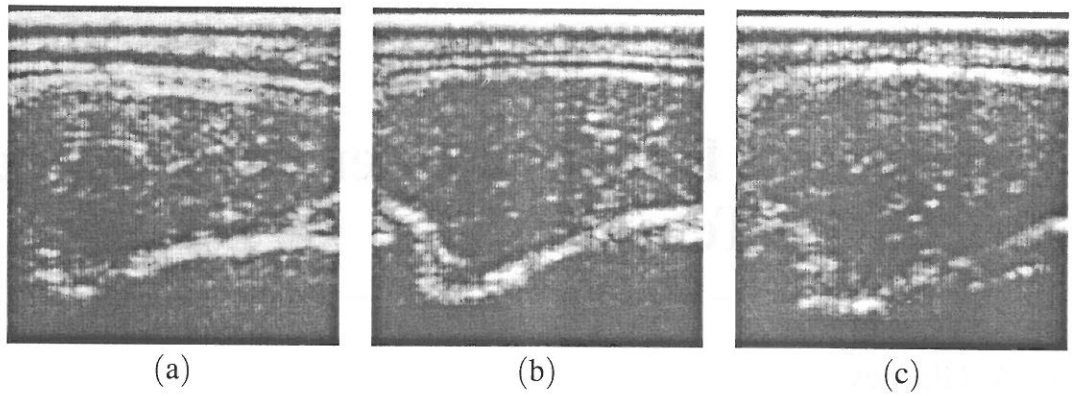


Fig. 1. Ultrasound images of the backs of three sheep.

magnetic resonance imaging (MRI) are ubiquitous in hospitals. They are also increasingly used in animal science and production to predict body composition (Simm, 1992). In particular, ultrasound has become widely used in animal breeding to assist in selection for leaner animals, because it uses cheaper and more-mobile imaging equipment than either X-ray CT or MRI. Unfortunately, ultrasound images are far more susceptible to noise and are therefore far harder to interpret, as can be seen in Figure 1. This is particularly true in this application, because we are using an old instrument which lacks a video output. The instrument operates by sending a pulse of sound waves of very high frequency into a subject (a 'squeak', about 250 times the highest frequency audible to the human ear). When the sound wave meets a boundary between two tissues, partial reflection occurs. The greater the difference in acoustic impedance between the two tissues, the greater the reflection. For example, more energy is reflected from a muscle-bone interface than from a muscle-fat one. This faint echo is received by the instrument and converted into electrical signals which are displayed on a video monitor, with time delay interpreted as depth.

A Vetscan ultrasonic instrument was used to scan the backs (at the 13th thoracic vertebra) of each of 72 Suffolk ewe lambs in two consecutive weeks (Glasbey *et al.*, 1996). A subset of the images, the pairs from 24 sheep, were interpreted manually, by using a mouse to control a screen cursor and identify 4 tissue boundaries. For example, Figure 2 shows manually-identified boundaries for the images in Figure 1. The topmost piecewise-linear line locates the boundary between the skin (above) and the fat (below). The piecewise-linear line below this identifies the fat-muscle boundary. Finally, the two lowest lines, which are piecewise-quadratic, show the top and bottom of a rib and, on the left of each image, part of the backbone.

The aim of the image interpretation is to estimate the average fat depth and maximum muscle depth for each sheep, from which whole body composition can be estimated. Obtaining these measurements manually is a slow and tedious task, and the results will vary between scientists. It is desirable to fully automate the analysis. Some success has been achieved with low-level image processing algorithms (Glasbey *et al.*, 1996), but these are unlikely to generalise easily to other scanners, sheep breeds or scanning positions. Trainable models have been applied in the interpretation of human echocardiograms (Baldock, 1992). Therefore, the segmentation problem looks well suited for stochastic templates, whereby high-level knowledge on anatomical structure can be incorporated into a segmentation algorithm.

Bayesian Approach

The templates can be summarised by 15 landmark points, which are also shown in Figure 2. The top 10 points specify the top and bottom of the fat boundary, while the bottom 5 points specify the rib and backbone. In the latter case, quadratic interpolation has been used and the rib is constrained to a fixed width. Sixteen parameters are required to specify the 15 landmarks,

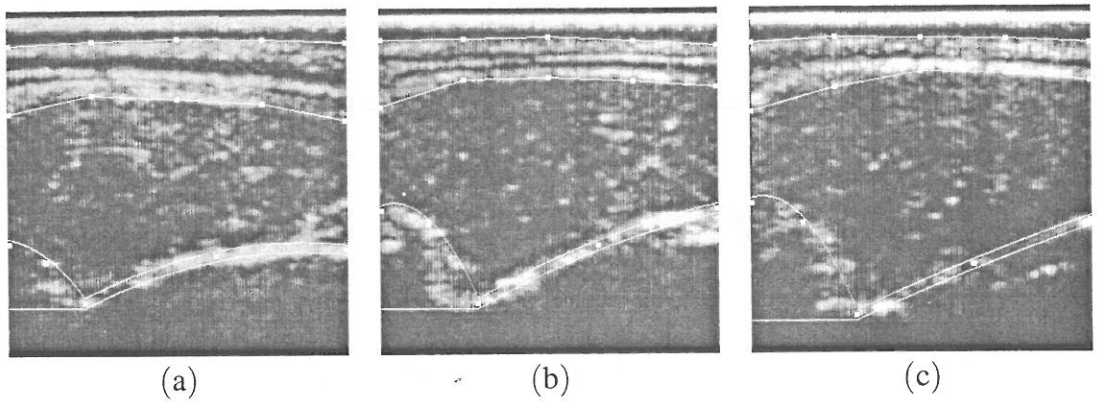


Fig. 2. Ultrasound images from Figure 1 with manually-identified boundaries superimposed.

namely 15 row locations, together with the column location where the rib joins the backbone. The remaining 14 column locations are fixed, or defined relative to this single column location.

We assessed the 16 parameters from the 48 training images for multivariate normality. Figure 3 shows normal probability plots for the first 4 principal components (see, for example, Krzanowski, 1988), which account for 91% of the variability. The linearity of these plots supports the assumption of normality. Similarly, the pairwise scatter plots in Figure 4 are consistent with normality. Therefore, the prior probability can be specified by

$$P(\beta) = \frac{1}{(2\pi)^{m/2}} \exp\left[-\frac{1}{2}\beta^T\beta\right],$$

where β denotes the vector of the first m principal components. Figure 5 shows the effects of the first 4 principal components on the template. As can be seen, the first principal component mainly influences the fat depth, the second one specifies the position of the scan relative to the backbone, the third one changes rib angle and the fourth makes more subtle changes to the fat depth and rib position.

If the top-most black and light-grey bands in the images (the first 18 rows of each image, which show the transducer-skin boundary) are ignored, the templates in Figure 2 partition the images into 5 segments, corresponding to: (1) skin, (2) fat, (3) muscle, (4) bone and (5) internal organs. For simplicity, pixels in segment k are assumed to be independently, identically, normally distributed with mean μ_k, standard deviation σ_k. Therefore, the likelihood of the image, conditional on a template specified by β, is given by

$$P(y|\beta) = \prod_i \prod_j \frac{1}{\sqrt{2\pi\sigma^2_{f_{i,j}}}} \times \exp\left[-\frac{1}{2\sigma^2_{f_{i,j}}}(y_{i,j} - \mu_{f_{i,j}})^2\right],$$

where '|' denotes 'conditional on', $y_{i,j}$ is the pixel value in row i column j of the image and $f_{i,j}$ denotes its segment label (an integer between 1 and 5), which is a function of β, but, to prevent the notation from becoming too complicated, this dependence has not been indicated. The independence assumption is clearly inappropriate, because correlations between adjacent pixels are evident in Figure 1, but has been used by others (Phillips and Smith, 1994), so we will start by using it here. Alternatively, we could have tried using the ASM methodology (Cootes *et al.*, 1995) to align edges, but the presence of multiple and very noisy edges in the images made us favour a statistical approach. A benefit of this model is that $P(y|\beta)$ is fast to compute, using

$$P(y|\beta) = \prod_{k=1}^{5} \prod_j \frac{1}{(2\pi\sigma_k^2)^{(g_{k,j}-g_{k-1,j})/2}} \times \exp\left[-\frac{1}{2\sigma_k^2}\left\{(A_{g_{k,j},j} - A_{g_{k-1,j},j}) - 2(B_{g_{k,j},j} - B_{g_{k-1,j},j})\mu_k + (g_{k,j} - g_{k-1,j})\mu_k^2\right\}\right],$$

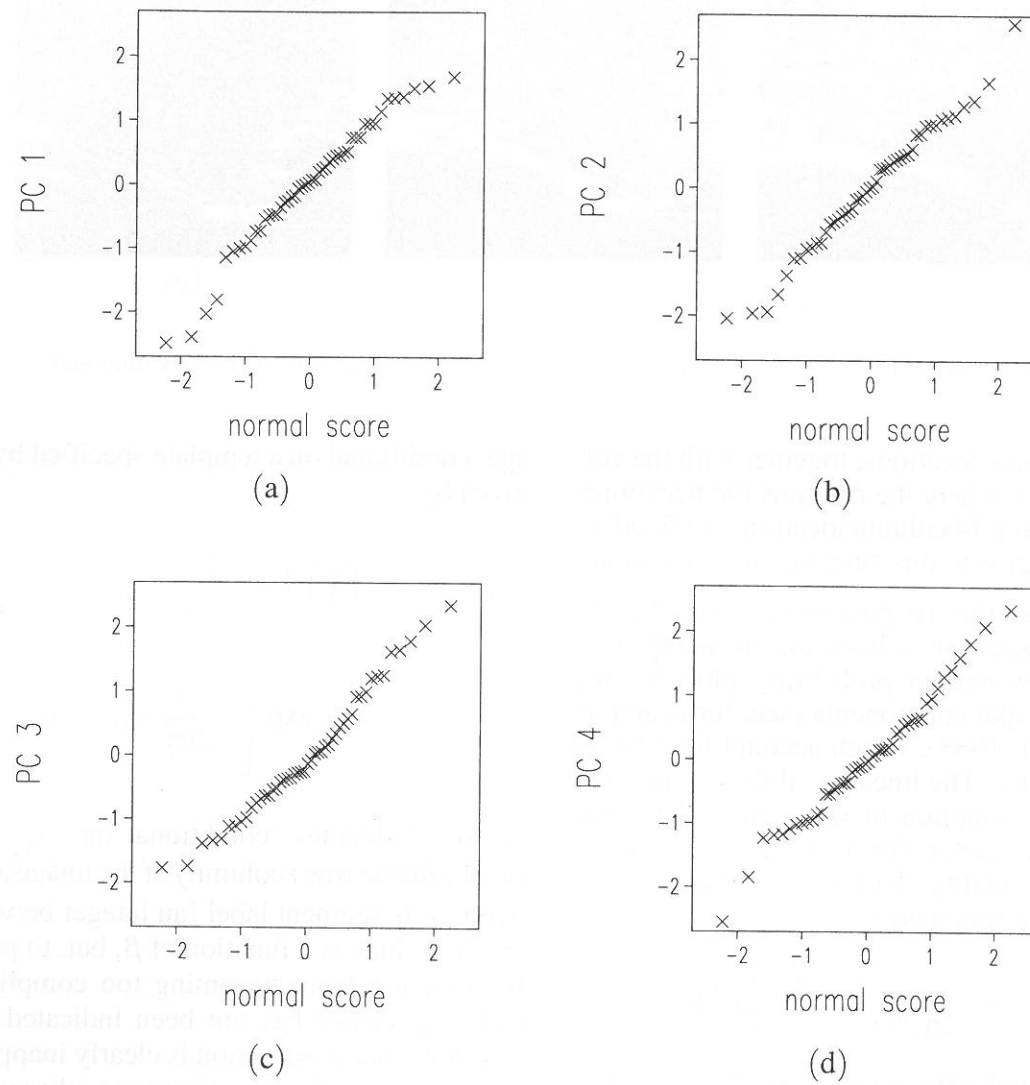


Fig. 3. Normal probability plots of first four principal component scores for 48 training images.

where $g_{k,j}$ denotes the last row of segment k in column j, $g_{0,j}$ is set to 18,

$$A_{h,j} = \sum_{i=0}^{h} y_{i,j}^2 \quad \text{and} \quad B_{h,j} = \sum_{i=0}^{h} y_{i,j} \ ,$$

which can be computed recursively down each column, by

$$A_{h,j} = A_{h-1,j} + y_{h,j}^2 \quad \text{and} \quad B_{h,j} = B_{h-1,j} + y_{h,j} \ .$$

Parameters μ_k and σ_k can be estimated from the 48 training images, as the sample means and standard deviations of pixel values $y_{i,j}^{(u)}$ for which $f_{i,j}^{(u)} = k$, where $u = 1 \ldots 48$ indexes the 48 images. This is equivalent to maximising $\prod_u P(y^{(u)}|\beta)$. These estimates are given in the first row of Table 1. All images were standardised to have the same mean intensity, and pixel values range from 0 to 255, with 0 corresponding to black and 255 to white. As we would expect from inspection of the images, segment 2 (the fat layer) has the highest mean pixel value, and segments 2 and 4 (fat and bone) have the largest standard deviations.

Bayes' theorem gives the posterior probability for the template, conditional on an image, proportional to the product of the prior probability

	μ_1	μ_2	μ_3	μ_4	μ_5	σ_1	σ_2	σ_3	σ_4	σ_5
estimates	77	120	31	65	20	57	64	26	63	16
modified values	−68	2	−20	84	19	49	67	45	99	27

Table 1. Values of parameters in likelihood.

and the likelihood, that is,

$$P(\beta|y) = \frac{P(\beta)P(y|\beta)}{P(y)}.$$

(See, for example, O'Hagan, 1994.) The divisor, $P(y)$, is a normalising term which ensures that $P(\beta|y)$ integrates to unity. Because $P(y)$ is a constant we can ignore it when maximising $P(\beta|y)$. It would have been possible to place prior distributions on μ and σ, but we found it to be simpler and as effective to keep them fixed. The Bayesian framework also gives us the option to explore the full posterior distribution to assess the range of templates compatible with an image, but at present we have concentrated solely on the distribution mode.

We maximised $P(\beta|y^{(u)})$ for each of the 48 images in the training set, using the Nelder-Mead optimisation algorithm (NAG, 1993) from 20 starting positions, each being the best of 100 randomly chosen values from the prior distribution for β. We also experimented with alternative optimisation methods, in particular the use of Markov Chain Monte Carlo (Phillips and Smith, 1994), but found them to be less effective for finding the mode than for studying the posterior distribution. The results we present are based on using $m = 4$ principal components. Very similar results were obtained using more components, but would not extend so readily to the approach taken in §4 and so they are not given here. Although the manual template was approximately recovered in most cases, this was not consistently the case. Figure 6 shows the results for 3 images. In particular, it can be seen in Figure 6(b) that the rib has been misplaced. The algorithm also had trouble locating the backbone, but we are less surprised by this, because information on the backbone is very sparse in the ultrasound images. Overall, fat depths estimated by the algorithm agreed well with those from the manual template, with a sample correlation coefficient of 93%, but muscle depths agree less well, with a correlation of 24%.

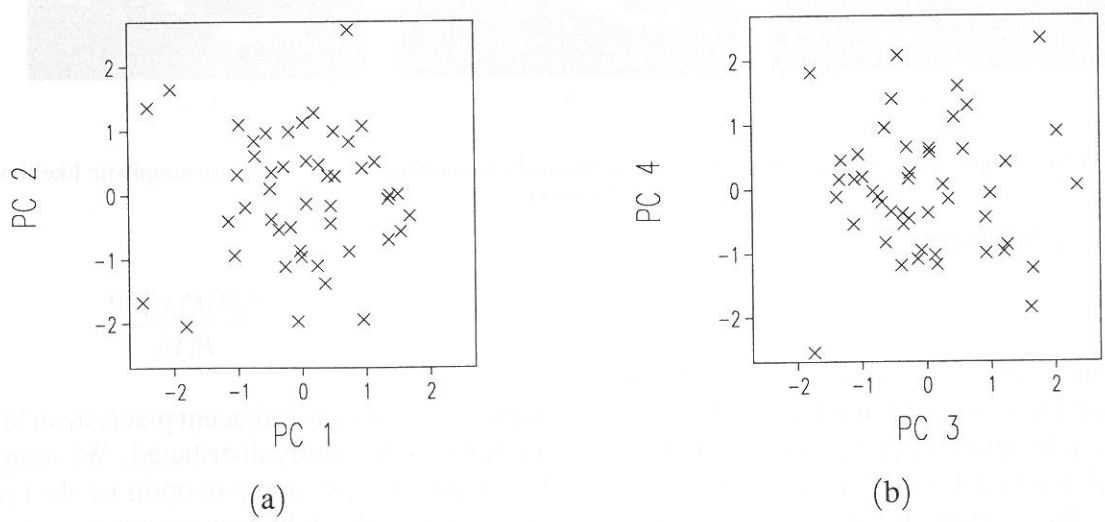


Fig. 4. Scatter plot of principal component scores for 48 training images: (a) first v. second, (b) third v. fourth.

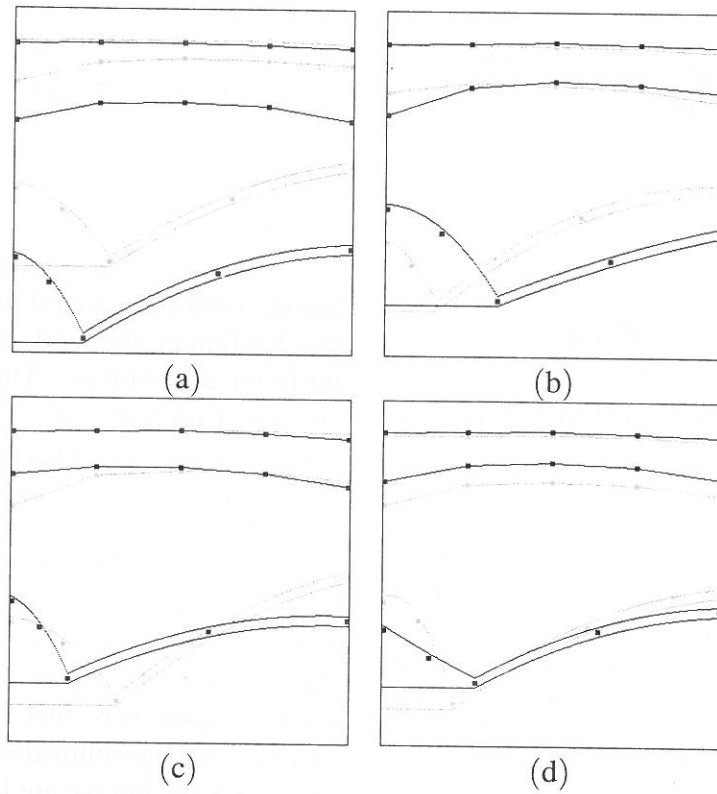


Fig. 5. Average template + 2 standard deviations (solid line) and average template − 2 standard deviations (dotted line) for: (a) first, (b) second, (c) third and (d) fourth principal component.

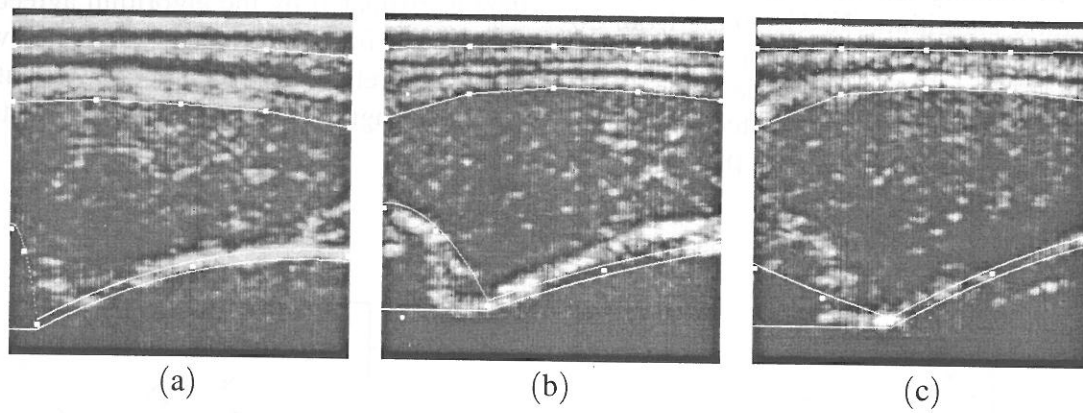


Fig. 6. Ultrasound images from Figure 1, together with poorly located boundaries based on simplistic likelihood function.

Modified Likelihood

The main reason for the poor results in the previous section is that the likelihood, $P(y|\beta)$, is only a crude approximation to the distribution of y, and it dominates $P(\beta)$ in the expression for $P(\beta|y)$. We found that the templates were recovered more consistently if the likelihood was downweighted by using

$$P(\beta|y) = \frac{P(\beta)P(y|\beta)^{\gamma}}{P(y)},$$

with $\gamma \ll 1$, because adjacent pixels in an image are not independently distributed. We searched for values of μ_k, σ_k and γ to optimise the recovery of the templates, by maximising the criterion

$$\frac{\bar{X}}{\sqrt{\frac{1}{47}\sum_u (X^{(u)} - \bar{X})^2}},$$

where u indexes the 48 training images,

$$X^{(u)} = \log P(\beta^{(u)}|y^{(u)}) - \max_{\beta' \in S^{(u)}} \left[\log P(\beta'|y^{(u)})\right],$$

$$\bar{X} = \frac{1}{48}\sum_u X^{(u)},$$

and $S^{(u)}$ is the set of integer values of β' within 3 standard deviations of the mean template, but excluding those values near $\beta^{(u)}$, i.e.

$$\{\beta' \ : \ \beta'_i = \pm 3, \pm 2, \pm 1, 0 \ \text{ for } \ i = 1 \ldots 4; \ \max_i |\beta'_i - \beta_i^{(u)}| \geq 1\},$$

which has $7^4 - 2^4 = 2385$ elements. Alternatively, randomly-selected values of β' could have been used. Ideally, we would like all the $X^{(u)}$ to be positive, because then $P(\beta^{(u)}|y^{(u)})$ would be larger than $P(\beta'|y^{(u)})$ for any other value of β'. The criterion we used is somewhat *ad hoc*, but is a differentiable measure of positivity of the X's, which is invariant to multiplying factors in the probabilities.

The Nelder-Mead optimisation algorithm was used from multiple starting points. The maximum value found for the criterion was 1.01, which, if the X's are approximately normally distributed, implies that in 16% of cases X will be negative. For comparison, the criterion was only -0.26 for the original set of parameters. The new values are given in the second row of Table 1, and $\gamma = 1/1700$. These values can be seen to be quite different from the first set, and to make no physical sense because pixels do not take negative values. Therefore, the likelihood should be regarded as an empirically-derived cost function. We also experimented with other functional forms for the likelihood, but found none better, and so retained the original form because of its computational convenience.

The choice of value for γ can be understood by considering the dependence structure in images. Examination of one image indicated

$$\text{cov}(z_{i,j}, z_{k,l}) = 0.92^{|i-k|} \times 0.988^{|j-l|},$$

where $z_{i,j}$ is the (i,j)th pixel value, after standardisation to zero mean, unit variance, i.e.

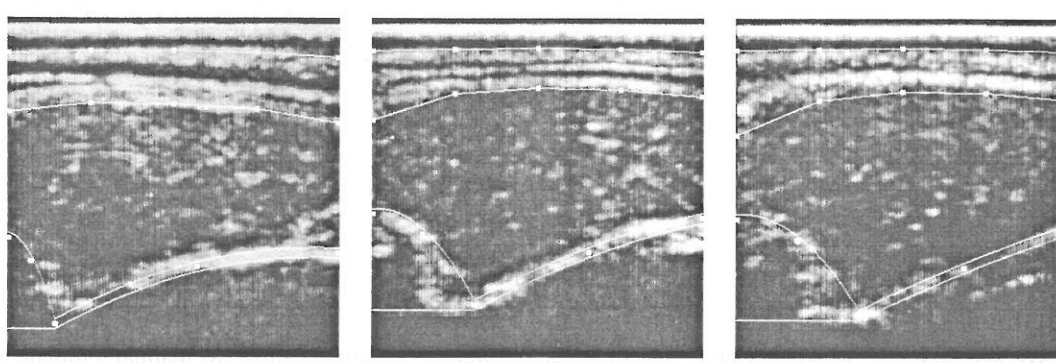
$$z_{i,j} = \frac{y_{i,j} - \mu_{f_{i,j}}}{\sigma_{f_{i,j}}}.$$

The effective degrees of freedom of z is

$$\nu = \frac{2[\text{E}(\sum\sum z_{i,j}^2)]^2}{\text{var}(\sum\sum z_{i,j}^2)} = \frac{n^2}{\sum_i \sum_j \sum_k \sum_l 0.92^{|i-k|} 0.988^{|j-l|}},$$

obtained by approximating $\sum\sum z_{i,j}^2$ by a χ_ν^2-distribution, where n is the number of pixels in an image. An appropriate value of γ is ν/n. For example, if the observations are independent, then $\nu = n$ and $\gamma = 1$, whereas if all observations are perfectly correlated, $\nu = 1$ and $\gamma = 1/n$. The above correlation structure indicates $\gamma = 1/3000$.

The posterior probability based on the modified likelihood was again maximised for each of the 48 images in the training set, using the Nelder-Mead optimisation algorithm. The results are



(a) (b) (c)

Fig. 7. Ultrasound images from Figure 1, together with boundaries based on modified likelihood function.

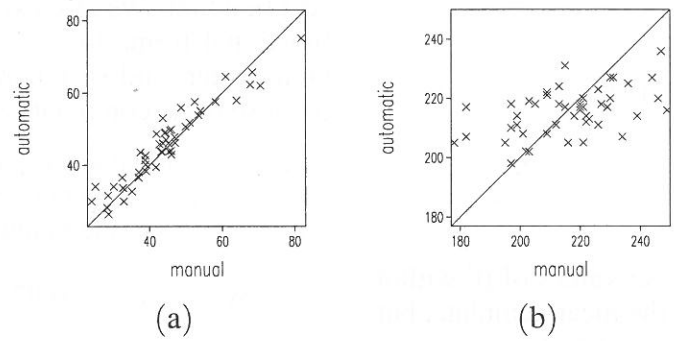


Fig. 8. Depths identified by algorithm plotted against those in manual template, for the training data: (a) average fat depth, (b) maximum muscle depth.

much improved, as can be seen in Figure 7, although they are less than perfect: the maximum muscle depth is overestimated in Figure 7(a), as is the fat depth in Figure 7(c). Agreement between manual and automatic results is better than before: the correlation is 95% for estimates of fat depths (Figure 8(a)), and 52% for muscle depths (Figure 8(b)). Note that correlations were not calculated on a leave-one-out basis, which would have involved recomputing the modified likelihood parameters 48 times.

For these data we do not know the true depths, but we can assess the accuracy of the methods by comparing the results for weeks 1 and 2 for the same sheep. Fat and muscle depths will be almost identical in the two scans, except for a small amount of growth in the intervening week. Figures 9(a) and (b) show estimated fat depth for the manual and automatic methods. It can be seen that the manual results are more consistent, with a correlation between weeks of 97%, while the automatic method has a correlation of 91%. The situation is reversed for muscle depth, shown in Figures 10(a) and (b). Here the automatic method, with a correlation between weeks of 72%, is far superior to the 24% for the manual method. We therefore conclude that manual estimation of muscle depth is inconsistent, and that more of the inaccuracies in Figure 8(b) are due to the manual method.

The algorithm was applied to a validation data set consisting of pairs of images from a further 48 sheep. As is to be expected, the fit is not quite so good as for the training data. Figure 9(c) shows estimates of fat depth for weeks 1 v 2, for which the correlation is 65%. Figure 10(c) shows muscle depth, with a correlation of 58% which is still larger than the 24% produced by the manual method on the training data. Figure 11 shows the two sheep for which

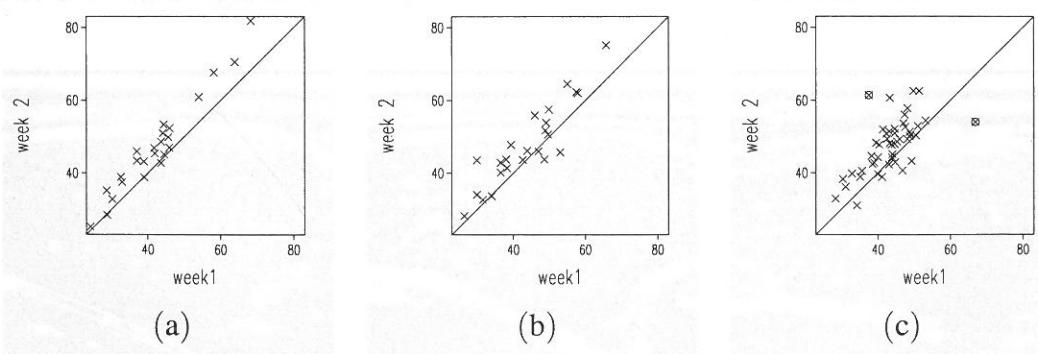


Fig. 9. Average fat depth for week 1 plotted against that for week 2, obtained using: (a) manual template applied to training data, (b) algorithm applied to training data (c) algorithm applied to validation data (images corresponding to circled points are displayed in Figure 11).

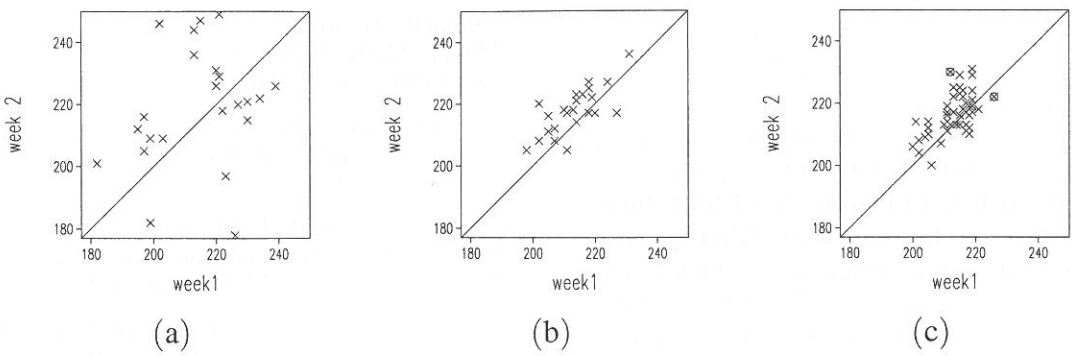


Fig. 10. Maximum muscle depth for week 1 plotted against that for week 2, obtained using: (a) manual template applied to training data, (b) algorithm applied to training data (c) algorithm applied to validation data (images corresponding to circled points are displayed in Figure 11).

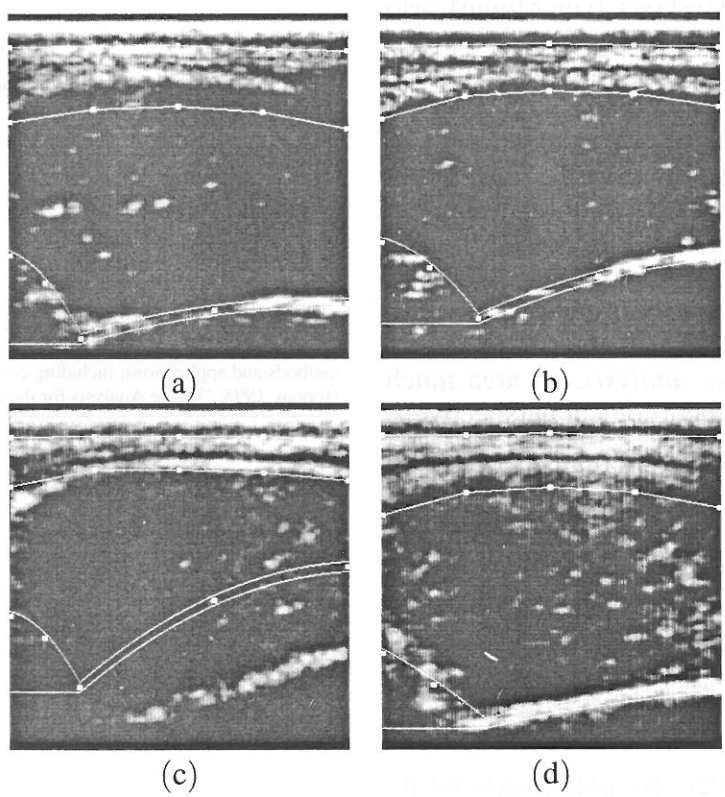


Fig. 11. Two sheep in validation data set for which worst results produced, as indicated by circled points in Figures 9 and 10: sheep 1, (a) week 1, (b) week 2; sheep 2, (c) week 1, (d) week 2.

the worst results were obtained, indicated by the circled points in Figures 9(c) and 10(c). It can be seen that, in Figures 11(a) and (d) the fat depth is overestimated, whereas in Figure 11(c) the muscle depth is underestimated. It is not apparent why these errors occurred, so it is difficult to see how to further modify the method in order to overcome them. We conclude that, overall, the results are encouraging, especially when the poor quality of the original images are borne in mind: estimates of sheep fat and muscle depths are comparable in accuracy with manual interpretation of images.

Discussion

We have demonstrated how training data can be used in a Bayesian approach to identify not only the prior distribution but also the likelihood function such that the mode of the posterior density coincides with the template. Thus, we have maximised the amount of information obtained from the training data. Alternative approaches would be to repeatedly return to each training image and modify the fitting strategy until all images are correctly segmented, to attempt to model the distributions in the training data more systematically, or to develop a mechanistic model taking account of animal physiology and ultrasound technology.

Further work is needed to explore alternative fitting strategies using Markov Chain Monte Carlo methods to estimate the posterior distribution. Experience is needed in applying the method to sheep of other breeds and to scans obtained at other anatomical locations, such as the lumbar region where the fat boundary has a different shape and ribs are absent. We will develop the methodology, and extend it to handle hierarchical structures (Phillips and Smith, 1994), to segment X-ray CT images of sheep. We also intend to place this work in the wider context of *semi-automatic image analysis*, an area much neglected from a methodological perspective.

Acknowledgements I am grateful to Geoff Simm for permission to use the ultrasound images. The work was supported by funds from the Scottish Office Agriculture, Environment and Fisheries Department.

References

[1] R. A. BALDOCK (1992) Trainable models for the interpretation of biomedical images, Image and Vision Computing, 10, 444–450.

[2] T. F. COOTES, C. J. TAYLOR, D. H. COOPER, J. GRAHAM (1995) Active shape models – their training and application, Computer Vision and Image Understanding, 61, 38–59.

[3] C. A. GLASBEY, I. ABDALLA, G. SIMM (1996) Towards automatic interpretation of sheep ultrasound scans, Animal Science, 62, 309–315.

[4] U. GRENANDER, M. I. MILLER (1994) Representations of knowledge in complex systems (with discussion), Journal of the Royal Statistical Society, Series B, 56, 549–603.

[5] W. J. KRZANOWSKI (1988) Principles of Multivariate Analysis: A User's Perspective, Clarendon, Oxford.

[6] NUMERICAL ALGORITHMS GROUP (1993) Library Manual Mark 16. NAG Central Office, 256 Banbury Road, Oxford OX2 7DE, UK.

[7] A. O'HAGAN (1994) Kendall's Advanced Theory of Statistics: Volume 2B, Bayesian Inference, Edward Arnold, London.

[8] D. B. PHILLIPS, A. F. M. SMITH (1994) Bayesian faces via hierarchical template modelling, Journal of the American Statistical Association, 89, 1151–1163.

[9] G. SIMM (1992) Selection for lean meat production in sheep, In Recent Advances in Sheep and Goat Research, (A. W. Speedy, Ed.), pp. 193–215, CAB International.

Received: April, 1997
Revised: May, 1998
Accepted: May, 1998

Contact address:
C. A. Glasbey
Biomathematics and Statistics Scotland
JCMB, King's Buildings
Edinburgh EH9 3JZ
Scotland
phone: (44) +131 650 4899
fax: (44) +131 650 4901
email: chris@bioss.sari.ac.uk

C. A. GLASBEY is Head of Research in Biomathematics and Statistics Scotland. He has 20 years experience in statistical and image analysis methods and applications, including co-authoring a book (Glasbey and Horgan, 1995, "Image Analysis for the Biological Science", Wiley).