A Support Vector Machine for 3D texture-based classification of bone in CT images

T. Shepherd [a] and D.C. Alexander [a] *

[a]Centre for Medical Image Computing, University College London, Gower Street, London WC1E 6BT.

Abstract. Bone regions in CT images contain a range of intensities and textures, making their classification a non-trivial task. We show that a Support Vector Machine can be used to distinguish between bone voxels and those from background biological material based on differences in both intensity and texture. The SVM is trained with ground-truth texture samples from several CT images to calculate a decision boundary separating bone from non-bone. Signed distance functions are then calculated for test samples, giving both an estimate of class label and a measure of the classification certainty. Classification performance is compared directly with Linear Discriminant Analysis implemented with information from intensity and local texture measures. Results show that the SVM is a strong discriminator of whole-bone from non-bone. As such we expect the method to be useful in a wide range of image segmentation applications, particularly in combination with more sophisticated segmentation algorithms such as active contours.

1 Introduction

Bone segmentation in CT images typically exploits the high X-ray attenuation of bone material and for visualisation in 2D a simple intensity thresholding may suffice in distinguishing bone from background (non-bone) material. However, a single bone varies in density and internal structure, causing image intensity or texture measures to vary widely. Figure (1) shows an axial slice of a CT image of the human ankle, demonstrating the range of intensities and textures and confirming that distributions of intensities from regions of bone and non-bone have significant overlap. The outer bone has apparent gaps due to the variation in density of cortical bone. The inner bone has irregular dark and light patches where calcified tissue intrudes spongy cancellous bone. These issues make segmentation of individual bones of the hands and feet a challenging problem. The segmentation of small joint bones in the wrists and ankles is necessary for modeling joint loading and kinematics [1]. Other applications include the identification of fractured bones in orthopedic surgery planning [2] and quantitative osteoporosis studies [3], where variable mineral density is a fundamental problem.

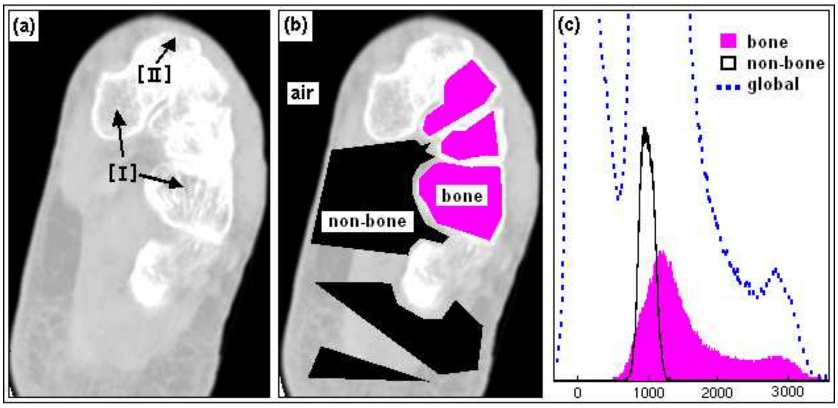

Figure 1. (a) Axial slice of a CT ankle. Arrows highlight examples of [I] the abundance of low intensity cancellous bone in inner-regions and [II] the variability of cortical bone thickness giving apparent discontinuity of the outer-bone surface. (b) Polygons manually drawn to define ground-truth regions of bone and non-bone. (c) Normalised histograms of CT intensity sampled from bone and non-bone regions across the whole of a 3D image. Intensities are in shifted Hounsfield Units as described in section (3). The global histogram is also shown, which includes voxels from air regions at the lowest intensities.

Bone segmentation methods such as active contour models (ACMs) exploit the intensity difference between bone and non-bone, but are confounded in regions of small joint bones where variability has a relatively large effect on the bone's form. Recent developments of ACMs for this application include the coupling of deformable models with region growing and region competition [1] and the formulation of new region energy terms [4]. In this paper we present

*E-mail: [T.Shepherd, D.Alexander]@cs.ucl.ac.uk

a supervised learning approach to 3D texture classification, which could similarly aid segmentation by deformable contour models. We use a Support Vector Machine (SVM) to characterise the distribution of 3D textures and intensities in each tissue type without the computation of local statistics. We hypothesise that a SVM operating on windows of raw intensities is able to classify bone better than statistical texture measures calculated over the same windows. We compare direct voxelwise classification performance using the SVM with that from using standard texture measures in classical Linear Discriminant Analysis (LDA).

2 Theory

2.1 Linear Discriminant Analysis

In LDA a linear decision boundary is calculated that separates two classes in the same space as the input feature vectors [5]. The training phase in LDA calculates a linear discriminant function

$$g(\boldsymbol{x}) = \sum_{i=1}^{d} w_i x_i, \tag{1}$$

i.e. a weighted sum of the components of the d-dimensional feature vector $\boldsymbol{x} = (x_1...x_d)^T$. A hyperplane is defined by constant $g(\boldsymbol{x})$. The task is to find the set of weights $\boldsymbol{w} = (w_1...w_d)$ and an offset from the origin in feature space, of the hyperplane which optimally separates two classes. After training, test data are classified by which side of this decision boundary they lie and a posterior probability of class affiliation is calculated from the perpendicular distance in feature space to the boundary.

2.2 The Support Vector Machine

The support vector method of machine learning was introduced by Cortes and Vapnik in 1995 [6]. A SVM maps input patterns into a higher dimensional space, using a kernel function, where it constructs a decision hyperplane. Maximally separated margins parallel to the hyperplane divide the new feature space into class-specific sub-spaces according to labelled training patterns. Often mapped patterns are not linearly separable, so that no hyperplane can separate two classes perfectly. In this case a cost is assigned to vectors lying between the hyperplane margins which, along with the positioning of margins, controls the optimization. A SVM is hence constructed by defining this cost, its input feature space and the kernel function. The results of training a SVM are the Lagrange multipliers used in optimisation, the data which lie on the hyperplane margins (support vectors) and the perpendicular distance of the hyperplane from the origin of the mapped feature space. New data is then classified by assigning to each pattern a decision-value d_{SVM} which is the perpendicular distance to the hyperplane. By convention the origin of the feature space is shifted to lie on the hyperplane, giving positive and negative decision-values so that sign indicates the predicted class.

2.3 Application to bone texture classification

We use 27-dimensional input patterns, comprising a list of intensities taken from each location in a $3 \times 3 \times 3$ cubic neighbourhood centred on a root voxel, as the input to a SVM. Figure (2) shows example intensity patterns from bone and non-bone regions and demonstrates that the intensity in bone regions varies more widely and the patterns themselves vary more in shape for bone regions than non-bone. Any recurring 3D texture captured by the cubic

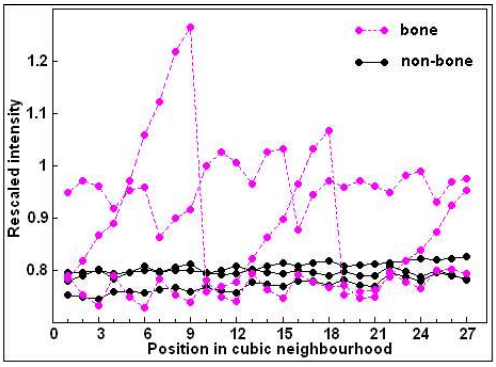

Figure 2. Three examples of 27-intensity patterns chosen at random from each of bone and non-bone regions. Intensity units are rescaled as described in section (4.3). Points are joined to show which intensities belong to a common pattern.

neighbourhood gives its own 'cluster' in input space. The number of clusters reflects the number of distinct textures within a region. By ordering the list of intensities by their position In the neighbourhood, this input feature space respects the spatial relationship between intensities in a neighbourhood, while such texture information is lost in the calculation of local statistics. We calculate four statistical texture measures which, along with CT intensity, give five features. We use the established texture features of local standard deviation, skewness, kurtosis and entropy [7, 8]. These are calculated over a $3 \times 3 \times 3$ cubic neighbourhood, centred on root voxels. LDA classification assigns two probabilities p_+ and p_- to each root voxel, i.e. the probability of belonging to each class. For convenience we convert these to a single signed decision-value p, using the formula:

$$d_{LDA} = p_+ - 0.5 + \left(0.5 \times sign[p_+ - 0.5]\right), \tag{2}$$

which allows the inferred class to be recognised by its sign and facilitates performance evaluation in the same framework as the SVM.

3 Data preparation

We use six CT images of human ankles. All images have in-slice pixel dimensions of 0.4 mm and original slice separation of 2 mm. We interpolate between slices to give 1 mm pseudo-slice separation. Intensities were originally in Hounsfield Units (HU), having a minimum at around -1000, but have been shifted into the positive domain as is common among the CT community. In addition we standardised the images to occupy a common dynamic range before the linear rescaling described in section (4.3). Each image histogram is given the same lower and upper limits. All values below a lower limit I_{low} represent air voxels that do not appear in training or testing samples and were set to this lower limit. The average maximum value over the six images was chosen as an upper limit I_{high}. All values above this value are reduced to I_{high} and for any image with a maximum below this value, the maximum value is increased to I_{high}. Any neighbourhoods containing data at this upper limit are not used in training or testing samples. By imposing these limits we make the method invariant to the presence of both air and highly attenuating foreign bodies such as metallic implants, should we wish to include new data sets in the future.

4 Methods

4.1 Classification and performance evaluation

We use a data analysis framework written in Matlab, which incorporates freely available SVM and LDA tools [9, 10]. Ground-truth locations are stored as co-ordinate arrays built up by drawing polygons on every slice as in figure (1 b). The polygons aim to collect enough examples of all the textures and intensities that belong to each tissue type, and are not required to capture the shape of a given region. To classify unseen test data, we threshold the signed distance to the separating hyperplane. To assess the performance of each classification technique, we compute ROC curves by varying the threshold on d_{SVM} and d_{LDA} and counting true and false positives compared to the ground-truth. The area under the ROC curve provides a measure of the degree of discrimination between texture classes from each method. This area A_{ROC} is unity for a perfect classification and has a lower limit of 0.5 for a complete failure to classify better than random label assignment. We also calculate the operator-point, where a threshold gives optimal classification for specified relative costs associated with false positives and false negatives. This is calculated for use in simple threshold segmentation.

4.2 Experiments

To test each classifier we label ground-truth regions in all six data sets and use a leave-one-out strategy to compute average ROC areas. For a chosen number of data, a random subset of the available ground-truths is taken from the test image, and one-fifth of this number are taken at random from each of the five training images. Permuting the configuration of training and testing images gives six ROC curves. We calculate the mean and standard deviation of classification performance from these six curves. To address the hypothesis that the SVM using raw intensities out-performs standard texture measures used in LDA, we use both methods to classify texture at the same ground-truth locations. To separate the comparison of classification techniques with that of the features used, we use each method with both windowed intensities and statistical texture measures. When using SVM and LDA with texture measures, we use the complete 5-dimensional feature vector. In the case of LDA we also perform separate classifications using 2-dimensional vectors of intensity plus each of the texture measures in turn. This gives insight into the relative discriminatory power of each feature.

4.3 Optimising the SVM

The SVM has various parameters that can be changed and optimised. We chose the Radial Basis Function (RBF) kernel for feature space mapping, which has a parameter γ that controls the kernel width. The cost C assigned to vectors lying between the hyperplane margins, is another flexible parameter. We used a grid-search algorithm to optimise the combination of parameters γ and C. In a scheme similar to that suggested by Hsu et al [11] we identified a local peak in classification performance on the 2D surface of γ and C pairs, at $\gamma = 8, C = 8$. The number of ground-truth patterns used in training was seen to affect classifier performance for numbers less than a few hundred, with little to be gained by using more than a few thousand. We use 5000 training patterns, and classify the same number of test data. Note that this means 1000 samples from each of the five training images in any one configuration. A SVM performs better on data that are rescaled to a small, decimal range [11]. After standardising to a common range we rescale this range to values between 0 and 2, as gave the best performance in trials.

5 Results and discussion

Table (1) gives the results of both classifiers used with the various combinations of intensity and texture features.

Classification Method	**Mean A_{ROC}**	**Standard Deviation**
27-intensity SVM	0.974 (i)	0.0129
5-feature SVM	0.961 (ii)	0.0138
27-intensity LDA	0.960 (iii)	0.0230
5-feature LDA	0.966 (iv)	0.0091
2-feature LDA combining intensity with : std. deviation	0.969 (v)	0.0173
: skewness	0.958 (vi)	0.0205
: kurtosis	0.952 (vii)	0.0267
: entropy	0.959 (viii)	0.0123

Table 1. Classification performances of the SVM and LDA

The 27-intensity SVM gives the highest ROC area of 0.974 (i). However there is little difference between all performances and this classification accuracy is not significantly better ($t = 0.52$) than that of the next best method, namely the 2-feature LDA using intensity and standard deviation (v). Comparison of (i) and (iii) indicates that the technical differences between SVM and LDA, i.e. non-linear mapping and boundary-margin maximisation, improve the separability of intensity-patterns. Interestingly, the SVM technique does not aid the classification of statistical texture features, as indicated by (ii) and (iv). Of all methods using statistical texture measures, the LDA that uses intensity and local standard deviation gives the best classification accuracy (v). The remaining statistical features are the poorest measures of texture in bone and non-bone (vi)→(viii) and do not appear to complement standard deviation information in the LDA, although the difference between (iv) and (v) is not significant and may be due to random fluctuations.

The performance metric A_{ROC} only reflects the proportion of misclassification within a subset of data (ground-truth locations). For more qualitative results, we classify all non-air voxels throughout a 3D dataset and create binary images. Figure (3) shows an axial slice of a 3D image segmented by thresholding the results of some of the classifiers. The threshold chosen gives approximately equal numbers of false positives and false negatives in separate classifications of 5000 ground-truth samples taken from the same image. The 3D image segmented in figure (3) includes slices at the cuneiform region where bones are small and close together. The classification of ground-truth from this image gives ROC areas of 0.957, 0.954 and 0.938 for the 27-intensity SVM, the 2-feature LDA and 5-feature LDA respectively. One advantage of the SVM is revealed by less omission of internal bone in (a) than in (b) or (c). Morphological operations may improve this classification but would not help a more sophisticated segmentation algorithm. Both SVM and 2-feature LDA misclassify data in the narrow regions between close bones and at the skin/fat layer. These sub-classes of non-bone material may have relatively homogenous texture, like cortical bone. However, these regions were not well represented in training samples. Including some examples in the training set should help eliminate these false positives. Panel (c) indicates that, although the inclusion of skewness, kurtosis and entropy may reduce the separability of feature space, (reflected in the noisier background) this information enables the LDA to identify the skin/fat layer as non-bone. The binary nature of the SVM and LDA prohibits us from distinguishing more than 2 tissue types. However, active region models inherently perform binary classifications, so the SVM or LDA would be well suited for use in this type of segmentation [12, 13]. Remaining limitations of the SVM are the relative time complexity and training requirements. We found that a typical classification of 27-intensity or 2-feature vectors took around 40 times longer by the SVM method than LDA. The SVM method also requires larger training sets in order to infer rotation invariance.

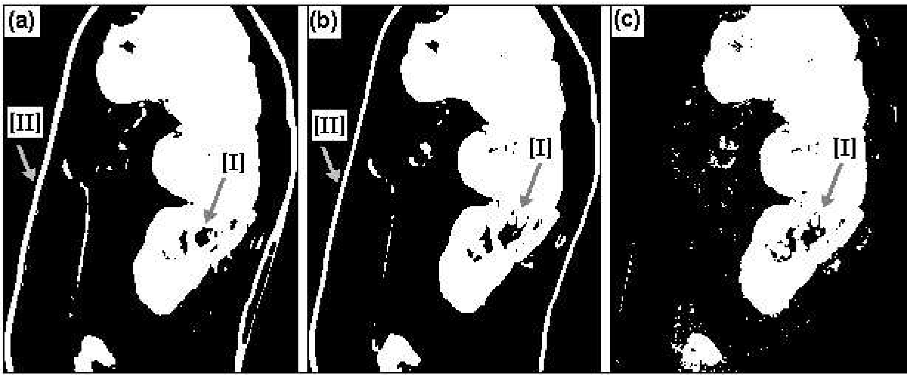

Figure 3. Slice of a 3D image showing bone (white) and non-bone (black) segmented by (a) the 27-intensity SVM, (b) the LDA that uses intensity and standard deviation and (c) the 5-feature LDA. Arrows show the misclassification of [I] internal bone regions and [II] the skin/fat layer.

However, the rotation invariant features used by the LDA do not appear to give this method an advantage for sample sizes above a few thousand.

6 Conclusions

The results show that a SVM can classify bone in CT images based on their varied 3D texture properties. The SVM groups many textures together in the same class, which enables more of the internal bone to be correctly classified. The kernel-based approach conserves the spatial relationship between intensities and removes the need for calculation of local texture features. The SVM gives a slight but not significant improvement over the traditional methods and both SVM and a LDA using statistical texture measures may be incorporated into more sophisticated segmentation algorithms. We intend to use the SVM to drive an active region model, which positions a contour based not only on edge properties but also by maximising a region measure calculated over the data it encloses.

Acknowledgements

This work was funded by an EPSRC studentship. We thank S.J.D. Prince for early reviews of the paper, S.Nedjati-Gilani for a practical introduction to SVMs, Depuy International for providing the images and M.Kiefte for correspondence regarding the use of the LDA toolbox.

References

1. T. Sebastian, H. Tek, J. Crisco et al. "Segmentation of carpal bones from 3d CT images using skeletally coupled deformable models." In *Proceedings of MICCAI'98: First International Conference*, pp. 1184–1194. Lecture Notes in Computer Science, Springer, Berlin, 1998.
2. S. Hirano & Y. Hata. "Fuzzy expert system for foot CT image segmentation." *Image and Vision Computing* **19**, pp. 207–216, 2001.
3. J. Sotoca, J. Iñesta & M. Belmonte. "Hand bone segmentation in radioabsorptiometry images for computerised bone mass assessment." *Computerized Medical Imaging and Graphics* **27**, pp. 459–467, 2003.
4. X. Pardo, M. Carreira, A. Mosquera et al. "A snake for CT image segmentation integrating region and edge information." *Image and Vision Computing* **19**, pp. 461–475, 2001.
5. R. Duda, P. Hart & D. Stork. *Pattern Classification (Second Edition)*. Wiley, *, 2001.
6. C. Cortes & V. Vapnik. "Support-vector networks." *Machine Learning* **20**, pp. 273–297, 1995.
7. T. Southard & K. Southard. "Detection of simulated osteoporosis in maxillae using radiographic texture analysis." *IEEE Transactions on Biomedical Engineering* **43**, pp. 123–132, 1996.
8. R. Haralick. "Statistical and structural approaches to texture." In *Proceedings of IEEE*, pp. 786–804. IEEE, 1979.
9. J. Ma, Y. Zhao, S. Ahalt et al. *OSU-SVM*. Support Vector Machine Toolbox, www.sourceforge.net/projects/svm.
10. M. Kiefte. *DISCRIM*. Linear Discriminant Analysis Toolbox, www.mathworks.com/matlabcentral/fileexchange.
11. C. Hsu, C. Chang & C. Lin. *A practical guide to support vector classification*. Technical Report, National Taiwan University, http://www.csie.ntu.edu.tw/ cjlin/papers/guide/guide.pdf, 2003.
12. J. Ivins & J. Porril. "Statistical snakes: Active region models." In *Proceedings of British Machine Vision Conference*, pp. 377–386. BMVA, 1994.
13. D. Alexander & B. Buxton. "Statistical modeling of colour data." *Int. Journal of Computer Vision* **44**, pp. 87–109, 2001.