Tina Memo No. 2007-005
Published in Proc. MIUA 2007

When Less is More: Improvements in Medical Image Segmentation through Spatial Sub-Sampling

P. A. Bromiley and N.A. Thacker

Last updated
13 / 4 / 2007

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

When Less is More: Improvements in Medical Image Segmentation through Spatial Sub-Sampling

P. A. Bromiley and N.A. Thacker
Dept. of Medical Biophysics
Imaging Science and Biomedical Engineering Division
Medical School, University of Manchester
Manchester, M13 9PT, UK
neil.thacker@man.ac.uk

Abstract

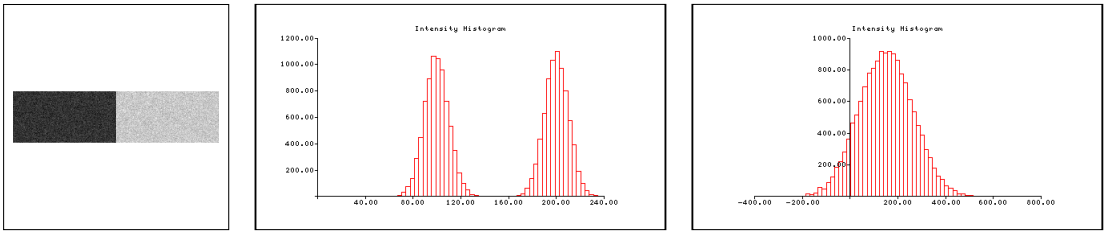
Segmentation is a common task in medical image analysis. It is frequently solved by fitting an intensity model, consisting of distributions for each pure tissue and each partial volume tissue combination, to the intensity histogram of the image data. However, this approach discards any spatial information present in the data. We present a method that recovers some of this information via regional sub-sampling during the fitting process. Experiments are performed on simple simulated data, simulated MR images from Brainweb, and real MR data from eight young normal subjects. The spatial sub-sampling procedure is shown to significantly improve the segmentation stability.

1 Introduction

One of the most common tasks in medical image analysis is segmentation: separating an image into individual tissue regions. This task is frequently performed (e.g. [9, 5]) by generating a model of the intensity distribution of the data, containing terms for each pure tissue and often terms for partial volume contributions (where a single voxel contains more than one pure tissue), and fitting this to the intensity histogram, using for example an Expectation-Maximisation (EM) algorithm [3]. However, this approach discards spatial information. Some researchers have investigated methods for recovering this information e.g. histograms built from local image phase [6] or Gaussian scale-space derivatives [4] have both been investigated in the context of image registration. In this work we propose using spatial maps of tissue locations to aid in the estimation of unstable parameters via a process of sub-sampling. The approach is similar to the use of spatial Bayesian priors, but bootstraps the maps from the data to avoid bias.

The segmentation algorithm used as a basis for the work described here has been presented previously at MIUA [11, 1, 8, 7, 10], and so we provide only a brief overview. The algorithm is based on a model of the intensity distribution for each tissue, consisting of pure tissue components represented by Gaussian distributions, and partial volume components (assumed uniform between pure tissue) represented for computational reasons by Gaussian distributions convolved with triangular distributions. Each pure tissue component contains parameters for the mean, the standard deviation, and a prior term representing frequency of occurrence, giving a total of four parameters. Unlike more conventional mixture models, the partial volume distributions are constrained by the parameters of the pure tissues they contain, which determine the mean and standard deviation, and therefore only contain only one additional free parameter, the frequency of occurrence. Partial volume can be set to zero for combinations of tissues that are known a-prior not to occur. This model is fitted to the intensity histogram of the data using the Expectation-Maximisation (EM) algorithm, providing a map of the volume of each tissue within each voxel.

Tissues tend to occur in contiguous blocks and for some sub-regions of the image pure tissue peaks can be easily identified. In principle the model we require is that which would provide a good fit to any sub-region of the image following estimation of the tissue normalisation parameters. Fitting a single combined histogram will not enforce this constraint, and the amount of information lost may be significant. We could attempt multiple simultaneous fits of disjoint regions, with common distribution parameters. Sampling would need to be done without replacement in order to avoid statistical over-fitting, and we can imagine that there will be some subsets which will maximally constrain some parameters. A simple experiment is provided here to illustrate the problem of regional contamination on the density fitting process. Logically, we would predict that if it were possible select a-priori all of the voxels containing one tissue this would provide the most accurate estimate of its parameters. For practical exploitation of this observation, we take a bootstrap approach, the aim being to generate sub-sampled data which consists primarily of one tissue type, providing a clearer peak to fit than is present in the histogram of



(a) (b) (c)

Figure 1: The synthetic data (a) with 10% noise, histogram at 10% noise (b) and histogram at 100% noise.

the data as a whole. We believe that this approach is justifiable within the context of an EM fitting algorithm, as changing the sample selected for the estimation of parameters should not introduce bias, provided the model is still valid for the sub-sample. The standard proof of convergence for the EM algorithm therefore extends to scenarios in which parameters are iteratively updated individually, but for a set of separate likelihood functions defined according to their own data sub-set rather than a likelihood defined for the entire data set. It is the theoretical freedom this provides, together with the regional coherence of data, which we exploit for improved parameter stability.

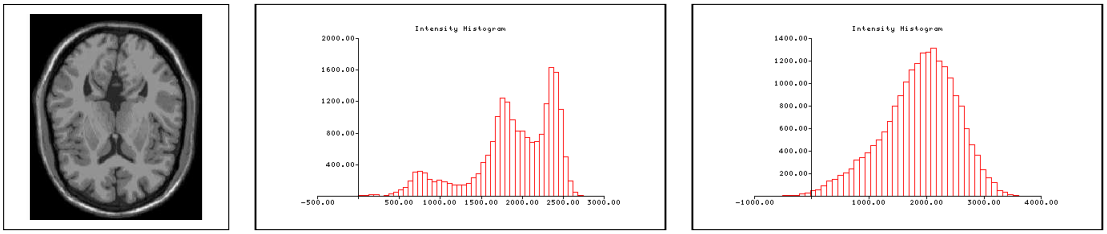
In order for sub-sampled regions to be suitable for parameter estimation they must satisfy the sampling requirements of the original density model. We can not therefore simply use a threshold on tissue probability to select data as this would impose hard cut-offs on the intensity histograms, which would not be described by the model. Instead we degrade an initial estimate of the tissue boundaries by blurring and quantisation in order to de-correlate the process of sampling with the grey-level values. The sample of a given tissue is therefore enhanced and close to a pure tissue sample, but the overall histogram distribution still conforms to the assumed model. The parameters of each pure tissue obtained from their individual sub-samples of data are then recombined to produce a complete tissue model, and the process iterated until convergence. Results from both Brainweb simulated MR images and real MR data, presented below, indicate that significant improvements in parameter stability can be gained through incorporating spatial information in this way.

2 Method

The mathematical formulation of the algorithm, and the parameter update equations, can be found in our previous publications [11, 1, 8, 7, 10]. The algorithm is capable of segmenting multi-dimensional data, and also of incorporating gradient information into the intensity histogram and model, although neither of these features were used here. In this work, we investigate the inclusion of spatial information into the above framework. At each step of the EM optimisation, the current tissue model can be used to generate tissue maps showing the volume of each pure tissue within each voxel. These can be binarised at the 50% level to produce maps of the spatial locations of each tissue type. These masks are blurred slightly by expanding each pixel to a 3x3 region, on the assumption that the main axes of the image do not align with any tissue boundary over the whole image plane and that over distances of a few pixels the tissue can switch entirely from one class to another. These masks are then used to sub-sample the original data set; tissue frequencies are maximised on each sub-sample to take account of the changing amount of data, and an expectation step is performed. The mean and standard deviation of each pure tissue is optimised from its respective sub-sample of the data. The parameters of each tissue are then recombined into a complete model, and estimation of the tissue frequencies is performed again prior to the next expectation step. The process is iterated until convergence.

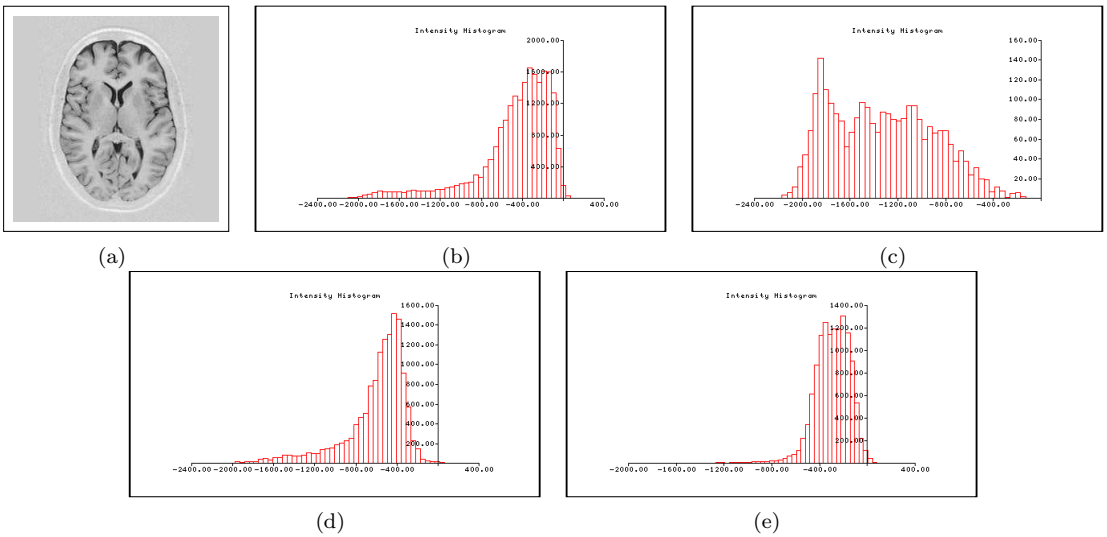
Three sets of experiments were performed in order to determine whether the incorporation of spatial information in the above framework improves the accuracy of the segmentation. In each set of experiments, a consistent starting point for the EM optimisation was enforced by using a hand-written initial tissue model, prepared through visual examination of the image histograms. 40 iterations of EM optimisation were applied in all segmentations; this was determined to be sufficient to ensure convergence to within 1% in all experiments.

In the first set of experiments, the synthetic data set shown in Fig. 1 was used, consisting of two blocks of uniform intensity, each 128 pixels wide by 64 pixels high, and having intensities of 100 and 200 grey-levels respectively (the darker region will be referred to as tissue 1, the brighter as tissue 2). 1000-iteration Monte-Carlo experiments were performed, in which independent Gaussian noise fields were added to the data prior to each iteration, and EM segmentation was performed, optimising only the parameters of tissue 1, holding the parameters of tissue 2



(a) (b) (c)

Figure 2: The Brainweb data (a), histogram at 2% noise (b) and histogram at 10% noise.



(a) (b) (c) (d) (e)

Figure 3: An example of one of the real MR data sets (a), its intensity histogram (b), and the spatially sub-sampled histograms for CSF (c), GM (d) and WM (e).

constant. Experiments were performed at each of ten levels of added noise, varying from $\sigma = 10$ to $\sigma = 100$ (i.e. from 10% to 100% of the separation in intensity between the two tissues. Sample histograms at the extremal noise levels are shown in Fig. 1. This procedure was then repeated a further 16 times, removing 8 columns from the right-hand side of the image each time, in order to simulate the effect of spatial sub-sampling of the data. Note that in the final experiment, the width of the tissue 2 region was reduced to 1 column rather than 0, to allow the use of the same, two-tissue, initial model. These experiments were intended as a form of calibration: they show the change in the accuracy of the optimised parameters for tissue 1 gained through reducing the contamination of the intensity histogram with tissue 2, for a range of different noise levels or, equivalently, separations between the mean intensities of the tissues. Note, however, that application of the spatial sub-sampling to real data would involve simultaneously masking-out and fitting each tissue. This experiment does not simulate the simultaneous optimisation of the parameters of tissue 2, which would introduce errors into the combined process, and thus under-estimates the errors on the tissue 1 parameters and also therfore, the gains achieved through application of the spatial sub-sampling to real MR data.

Simulated MR images obtained from Brainweb [2] were used in the second set of experiments. A simulated T1 volume with 0% inhomogeneity and 0% image noise, consisting of 1x1x1 mm voxels, was obtained. In order to reduce the processor time required, a single slice was selected from the data set, as shown in Fig. 2, and skull-stripping was applied using masks prepared from the Brainweb tissue phantoms, such that the segmentation was limited only to the cerebro-spinal fluid (CSF), white matter (WM) and grey-matter (GM). 100 iteration Monte-Carlo experiments were then performed, in which independent Gaussian noise fields were added to the data prior to each iteration, and EM segmentation was performed both with and without spatial sub-sampling of the data. Five such experiments were performed, with the standard deviation varying from 2-10% of the intensity of the brightest tissue, in steps of 2%. Coincidentally, these values are also approximately 20-100% of the separation between the GM and WM peaks. Sample histograms are shown in Fig. 2.

The final set of experiments used real IRTSE MR scans from eight young (34-46 years, mean age 40) normal subjects, each consisting of 1x1x3 mm voxels. For reasons of speed, single structure-rich slices from approximately

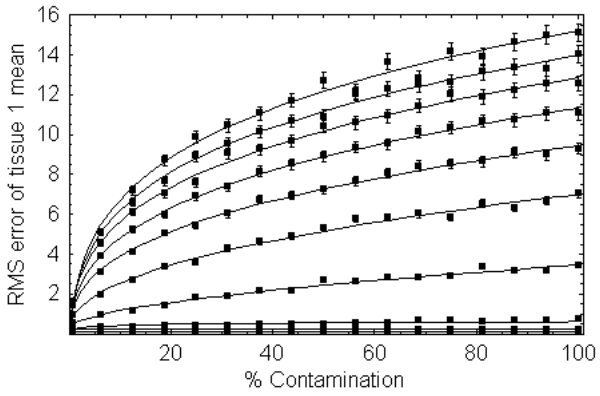
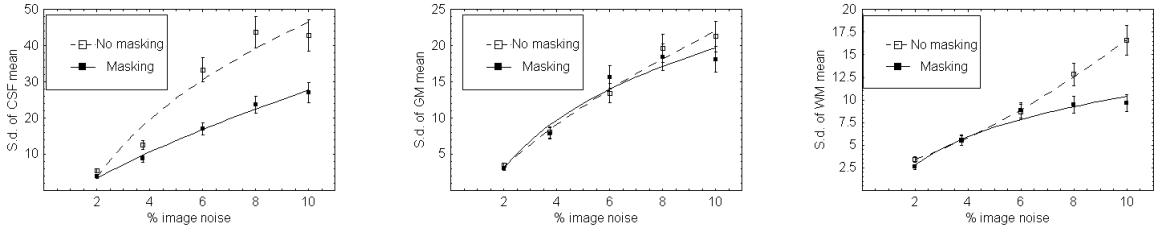
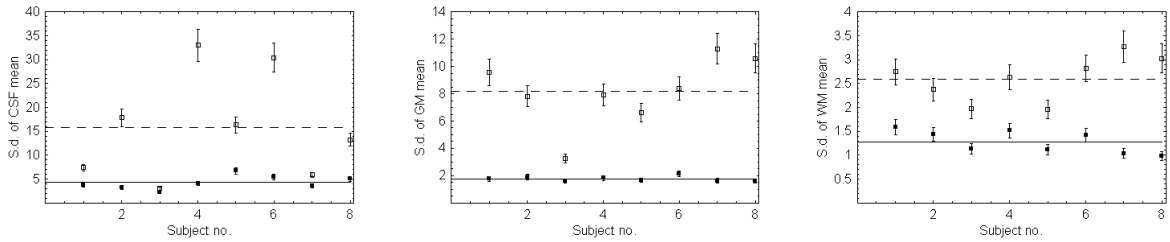


Figure 4: The RMS error on the tissue 1 mean in the experiments on synthetic data, plotted against the level of contamination of the intensity histogram with tissue 2. Power-law fits are shown. The lines are for 10-100% image noise, in steps of 10%, reading upwards from the bottom of the graph.



(a) (b) (c)

Figure 5: The standard deviations of the CSF (a), GM (b) and WM (c) mean intensities in a 100 iteration Monte-Carlo experiment on the Brainweb data, plotted against the total image noise (as a percentage of the dynamic range), for segmentations with and without the spatial sub-sampling of the data.



(a) (b) (c)

Figure 6: The standard deviations of the CSF (a), GM (b) and WM (c) mean intensities in a 100 iteration Monte-Carlo experiment on the real data, plotted against the subject no., for segmentations with and without the spatial sub-sampling of the data. The lines show the means across subjects, and the symbology is the same as in Fig. 5.

the same anatomical region were selected from each volume, and segmentation was limited to the CSF, GM and WM (located using hand-drawn binary masks). A sample slice and its intensity histogram are shown in Fig. 3. 100 iteration Monte-Carlo experiments were then performed, in which independent Gaussian noise fields with standard deviations equal to the that of the intrinsic image noise (estimated from the width of zero-crossings in horizontal and vertical gradient histograms) were added to the data prior to each iteration, and EM segmentation was performed both with and without spatial sub-sampling of the data. Fig. 3 shows example histograms produced by the spatial sub-sampling of the CSF, GM and WM at the start of the EM optimisation (i.e. prior to any optimisation of the hand-written model parameters).

3 Results

Fig. 4 shows the RMS error on the estimated mean intensity of tissue 1 in the experiments on simulated data. When the noise level is low, and so the tissues are well separated in the histogram, sub-sampling the data has no effect; this indicates that, at least on this example, the sub-sampling procedure does not introduce bias. However, at higher noise levels the spatial sub-sampling begins to have a marked effect: significant improvements in segmentation accuracy are gained. The results from the Brainweb data (Fig. 5) are broadly consistent; as the noise level is increased and the tissue intensities begin to overlap, sub-sampling significantly improves the stability of the segmentation (note the Brainweb does not provide mean intensity values for the simulated tissues, preventing any absolute tests of accuracy beyond e.g. simply measuring the volumes of segmented tissues or the degree of overlap of the segmentations with the phantoms). The results from the real data (Fig. 6) are significantly better than those obtained with the Brainweb data: whilst absolute segmentation accuracy cannot be measured without a gold-standard, parameter stability is improved by a factor of 2-4 through the application of spatial sub-sampling. The effect is greatest in the GM, since this is the most obscured tissue in the intensity histogram, forming a shoulder on the low-intensity side of the WM. Similarly, the effect is least marked in the WM, since this tissue forms a distinct peak in the histogram. The marked improvement of the sub-sampling performance on the real compared to the Brainweb data, particularly for the GM, may be due to sampling problems in the Brainweb simulation: spurious peaks were noted in the partial volume distributions in the histogram of the noise-free data, which are not consistent with typical partial volume distributions in real MR data.

4 Conclusion

Three main conclusions can be drawn from the results presented here. First, in situations where the intensity distributions of pure tissues overlap significantly, significant improvements in segmentation accuracy can be gained through recovering spatial information using the procedure described here. It should be noted that the initial tissue model, and thus the masks used to sub-sample the data, need not be highly accurate: hand-written models were used in the experiments, prepared simply from visual inspection of the intensity histograms.

Second, the spatial sub-sampling procedure discards large proportions of the data in the estimation of the tissue model parameters for each pure tissue, and furthermore the data samples change as the optimisation proceeds. Therefore, the results presented here demonstrate that the data sample can vary between the iterations of an EM optimisation without introducing biases or preventing convergence, as long as the sub-sampling procedure does not affect the goodness-of-fit of the model. This concept was adopted in our original segmentation algorithm, on which the current work is based, through the procedure of discarding the partial volume data in the parameter update step in order to reduce the sensitivity to the partial volume model used. The current work, using a much more aggressive sub-sampling of the data, demonstrates the general validity of this concept more fully. Furthermore, whilst not explored here, the potential computational benefits of sub-sampling the data and isolating specific model components for optimisation are clear.

Finally, it has been noted by many medical image analysis researchers that three of the main problems in the field, namely segmentation, registration, and shape modelling, can be viewed as three facets of an overall, unified problem, in that the results from any one of them can be used to perform the other two. The work presented here can be viewed as a first step towards the unification of segmentation and shape modelling: effectively it uses a crude, non-parametric shape model to sub-sample the data for segmentation. The improvements to segmentation accuracy gained indicate that further research in this direction could prove fruitful.

Acknowledgements

This work was supported by the the MIAS (Medical Images and Signals) IRC under EPSRC grant no. GR/N14248/01 and the MRC grant no. D2025/31. The software used in this study is freely available from www.tina-vision.net.

References

[1] P A Bromiley, N A Thacker, M L J Scott, M Pokrić, A J Lacey, and T F Cootes. Bayesian and non-Bayesian probabilistic models for medical image analysis. *Image and Vision Comput.*, 21(10):851–864, 2003.

[2] C A Cocosco, V Kollokian, R K-S Kwan, and A C Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4):S425, 1997.

[3] A P Dempster, N M Laird, and D B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Society*, 39:1–38, 1977.

[4] M Holden, L D Griffin, N Saeed, and D L G Hill. Multi-channel mutual information using scale space. In *Proceedings MICCAI'04*, pages 797–804, 2004.

[5] D H Laidlaw, K W Fleischer, and A H Barr. Partial-volume Bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Trans. Med. Imag.*, 17(1):74–86, 1998.

[6] M Mellor and J M Brady. Non-rigid multimodal image registration using local phase. In *Proceedings MICCAI'04*, pages 789–796, 2004.

[7] M Pokrić, N A Thacker, and A Jackson. The importance of partial voluming in multi-dimensional medical image segmentation. In *Proc. MICCAI*, pages 1293–1294, 2001.

[8] M Pokrić, N A Thacker, M L J Scott, and A Jackson. Multi-dimensional medical image segmentation with partial voluming. In *Proc. MIUA*, pages 77–80, 2001.

[9] P Santago and H D Gage. Statistical models of partial volume effect. *IEEE Trans. Med. Imag.*, 4:1531–1540, 1995.

[10] N A Thacker, M Pokrić, and D C Williamson. Noise filtering and testing illustrated using a multi-dimensional partial volume model of MR data. In *Proc. BMVC*, pages 909–919, Kingston, London, 2004.

[11] D C Williamson, N A Thacker, S R Williams, and M Pokrić. Partial volume tissue segmentation using grey-level gradient. In *Proc. MIUA*, pages 17–20, 2002.