Questions and Answers

Chapter 11

Questions

(Q1) *What are the advantages of atlas-based segmentation over other segmentation techniques?*

(Q2) *Why is the described non-rigid registration method superior to other techniques?*

(Q3) *What is the best value for the smoothness constraint weight of the non-rigid registration (section 11.3.4)?*

(Q4) *What if in section 11.4.2 the atlas most similar to the raw image were selected using the following criterion I invented: . . . ?*

(Q5) *When combining multiple atlas-based segmentations, what is the practical difference between NN interpolation with vote fusion and PVI with sum fusion?*

(Q6) *If the MUL atlas selection strategy is so much better than the others, then why is it not always used?*

(Q7) *How does atlas-based segmentation compare to manual segmentation?*

(Q8) *Are there parallels to the multiatlas segmentation method in pattern recognition?*

(Q9) *Could an active shape model be used as an atlas for segmentation?*

(Q10) *Why does the binary classifier performance model predict actual performance more accurately, yet the multilabel performance model gives better combined classification results?*

Answers

(A1) Which segmentation technique is the best strongly depends on the application. Atlas-based segmentation is most useful when the spatial location of a voxel, together with its gray value, determines which class it belongs to. Conversely, atlas-based segmentation should not be applied for problems where the classification depends purely on a voxel's gray value, for example when classifying different tissue types.

(A2) It is not in general superior. Again, the most suitable technique depends on the application. The reason for using this particular method is that it does not require single-modality image data. While strictly all images in the application presented here are confocal microscopy images, the imaging process causes very different gray value distributions for different individuals. The NMI image similarity measure is well-suited for this kind of situation, while other measures (Mean Squared Difference, Cross Correlation) are not.

(A3) Unfortunately, there is no *a priori* best value for this weight. There is always a trade-off between the smoothness and regularity of the transformation (higher weights), and good fitting of the data (lower weights). Choosing the best constraint weight is subject to active research, and is by no means a solved problem. For the bee brain microscopy images, we have found that the best results were achieved with a value of 0.1, but in other applications with different data and different objectives, the "best" values were very different. It is also possible to vary the constraint weight during the optimization to achieve better results.

(A4) Sorry, but—no! No criterion for the selection of the most similar atlas can produce segmentations of higher accuracy than a criterion that has knowledge of the true segmentation. This is precisely what the "Best SI" criterion is all about. So in our study at least, whatever the criterion, the SI values would never be better than using the criterion that selects based on the SI values themselves (which, let us mention it again, are not available in a real-world application anyway). Therefore, there is no criterion that would not still be outperformed by the MUL and even the AVG segmentation strategy.

(A5) By integrating the trilinear interpolation coefficients, PV interpolation leads to smoother transitions from one label to another and avoids jagged edges when interpolating oblique slices from an atlas. In general, we found NN and vote fusion to perform better for large numbers of segmentations (more than 30), since a distribution of samples from a large number of atlases approximates a continuous spatial distribution. PV with sum fusion seemed to perform slightly better for smaller numbers of segmentations. On a related note, due to the real-valued weights, it is much less likely to see a voxel tied between two or more labels using PV than using NN, so PV can help to reduce "rejected" classifications.

(A6) The MUL strategy has two drawbacks: it requires multiple atlases—a problem that it shares with the SIM and the AVG strategies. Furthermore, the raw image needs to be non-rigidly registered separately to each atlas, which again is also true for some criteria that could be used in a SIM strategy. Non-rigid registration is computationally expensive, so the bottom line is: for a MUL segmentation strategy, you need more atlases and more CPU time.

(A7) In short: we do not know. One important question is, how different is an automatic segmentation from a manual one, compared to the typical difference between two (or more) manual segmentations. The latter, the inter-observer difference, can only be determined by repeated manual segmentation of the same image. This is costly and time consuming and was therefore not done for the data we have used in this chapter.

(A8) A closely related technique in pattern recognition is the concept of "bagging" (short for bootstrap aggregation), a method for generating multiple, somewhat independent instantiations of a classifier by using different training sets. Multi-atlas segmentation is similar to this method in that it also generates multiple atlas-based classifiers using multiple training sets. In our case, the different training sets are the atlases, and the resulting classifiers are described by the coordinate transformations after registration of each atlas to the raw image.

(A9) Yes. In fact, it would be interesting to use a dynamic atlas during the registration stage rather than a static one. The dynamic atlas could be the average shape, with the most significant modes of variations added to it. The weights of all modes of variation could be incorporated as additional degrees of freedom into the registration process. So in some sense, this procedure would be similar to using an average shape atlas by itself, but with the additional benefit that the atlas could adapt to the individual that is being segmented. Another way to look at it is as a registration from two sides, where one image is deformed using a free-form deformation, whereas the deformation of the other, the atlas, is governed by the more specific ASM.

(A10) The multi-label performance model has substantially more free parameters than the binary model, and we only looked at one of them per segmented structure (i.e., sensitivity). While the binary model can only predict the probability of an error, the multi-label model actually makes a prediction of what the error will be. That is, the multi-label model allows us to model situations where, for example, misclassification of a voxel of class A into class B is much more likely than into class C. The binary model would only predict the likelihood of the misclassification, but would not provide any information as to the nature of the error.