Histogram-based Image Quality Checking

Hossein Ragheb
http://www.tina-vision.net/people.php
Neil Thacker
http://www.tina-vision.net/~nat
Paul Bromiley
http://www.tina-vision.net/~pab

ISBE, Faculty of Medicine
University of Manchester
Manchester, UK.

Abstract

Many medical image analysis algorithms make assumptions concerning the image formation process, the structure of the intensity histogram, or other statistical properties of the input data. Application of such algorithms to image data that do not fit these assumptions will produce unreliable results. This paper describes a technique for the automatic identification of images that do not have histogram structure consistent with that expected. The approach is based upon a component analysis followed by statistical testing. Experiments validate its use in the identification of quantisation problems and unexpected image structure. It is intended that this test will form one component of a quality control assessment, to aid in the use of sophisticated statistical image analysis software by non-expert users.

1 Introduction

Many complex image processing techniques, such as segmentation, registration and parametric image generation, have been shown to have utility in clinical applications. However, these techniques are always based on specific assumptions about the image formation process, the structure of the intensity histogram, or other statistical properties of the images. Considerable insight on the part of the end users may be required in order to avoid inappropriate application of such techniques to input data that do not fit these assumptions. Although a basic level of training with regard to loading data and executing analysis chains is common, it is generally not practical to provide adequate levels of training to end-users to enable them to assess the numerical or statistical stability of an algorithmic process on specific data. This can lead to inappropriate use of software and invalid research conclusions. Even the most commonly used packages, used in well funded studies, can be seen to have generated outputs which are quite clearly suspect [3]. To our knowledge there has been little effort expended towards solving such problems.

For CT and MR images, the DICOM header file may be used to check acquisition parameters such as temporal resolution, spatial resolution, weighting factors, and the presence or absence of contrast enhancements. We can also perform automatic data quality assessment prior to the main analysis (such as signal-to-noise checks [5]). However, such simple checks may not suffice to identify all possible image quality issues. In addition, the goal of automatic quality assessment software should be to provide end users with useful feedback and possible solutions when an input dataset fails a quality check.

Here, we use a histogram-based model of the data to ensure the valid use of statistical approaches. Specifically, we train the algorithm using a variety of compatible images. Our approach is based on fitting a combination of density functions to multiple independent subsamples of data. This model includes components for both pure tissue and partial volume voxels. Fitting parameters are updated using Bayes theory [1] which is used to estimate the components for an independent components analysis (ICA).

2 Algorithm

Training phase: The input image used for training (e.g. that shown in Fig. 1) is divided into J non-overlapping windows of equal size. This gives J different data histograms f_j $(j = 1,2,...,J)$ to which a unique histogram model is fitted. The model consists of I components $(i = 1,2,...,I)$ where each component is a density function $p(g|v_i)$ defined based on knowledge of the corresponding tissues. While the tissue parameters are identical for all histograms and are learnt through the optimisation of a global cost function, each histogram has specific weighting parameters α_{ij} which are updated using Bayes theory.

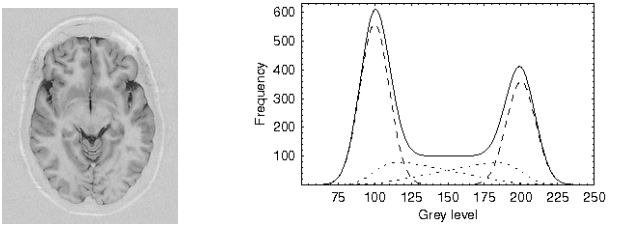

Figure 1: An example partial volume model for two pure tissues. Pure tissues have Gaussian distributions (dashed), while mixtures of tissues take form of triangular distributions convolved with a Gaussian (dotted). These are summed to give the overall distribution (solid).

The histograms are modelled using the approach equivalent to that described by Santago and Gage [6]. Their model consists of a delta function representing each pure tissue, and a uniform distribution between each pair of pure tissues that share a common boundary (see Fig. 1). Both types of distributions are convolved with a noise distribution which is assumed to be Gaussian. Therefore, pure tissues are represented by $(1/\sqrt{2\pi}\sigma)\exp[-(g-\mu)^2/2\sigma^2]$. We further refine the Santago-Gage model by splitting partial volume distributions into complementary pairs of triangular distributions, representing the volumetric contribution of each pure tissue to the partial volume voxel. If the triangular distribution is defined using the line equation $y = kx + c$, then its convolution with the Gaussian distribution is given by $\int_a^b (kt+c)(1/\sqrt{2\pi}\sigma)\exp[-(g-t)^2/2\sigma^2]dt$. Note that the mean parameter has no effect on the convolution process [3], and, the integral gives

$$-\frac{(kg+c)}{2}\{erf[\frac{g-b}{\sqrt{2}\sigma}]-erf[\frac{g-a}{\sqrt{2}\sigma}]\}-\frac{k\sigma}{\sqrt{2\pi}}\{\exp[-\frac{(g-b)^2}{2\sigma^2}]-\exp[-\frac{(g-a)^2}{2\sigma^2}]\} \quad (1)$$

The parameters a and b represent the non-zero range of the distribution. It is straightforward to find the intercept c and the slope k parameters of the line that defines the triangle. Absolute normalisation is not necessary at this stage and it is sufficient to assume that the maximum height of the distribution function is constant, or simply is equal to unity. Our

density functions which are represented by $p(g|v_i)$ are equivalent to ICA components. An example model consists of five Gaussians and eight corresponding triangular density functions between them. This makes four pairs of (a,b) together with an identical σ for all components. However, as parameter b for each range is identical to parameter a for the neighbouring range, six parameters are sufficient to account for all the model components. These are the five mean parameters of the five Gaussians plus the σ parameter. It is sufficient to set initial values to five equal partitions of the widest existing histogram range.

The next step is to determine all weighting parameters α_{ij} for histograms f_j and components $p(g|v_i)$ from the EM algorithm. We approximate our data histogram as a linear combination of all density f functions defined so that $f_j \approx \sum_i \{\alpha_{ij} p(g|v_i)\}$. The process of estimating the weighting parameters is iterative with $\alpha'_{ij} = \sum_g \{f_{gj} P(v_i|g)\}$. Probabilities are computed using the density functions and current weighting parameters α_{ij}. Specifically

$$P(v_i|g) = \alpha_{ij} p(g|v_i) / \sum_i \{\alpha_{ij} p(g|v_i)\} \tag{2}$$

The initial values used are $\alpha_{ij} = 1$. The equations are iterated until the parameters converge, when $\alpha'_{ij} \approx \alpha_{ij}$. Given α_{ij} it is straightforward to compute the cost function L_j for the histogram f_j. The appropriate cost function can be derived from the probability of getting the observed sample using Poisson assumptions. This results in the conventional likelihood function $L_j = -\sum_g \{f_{gj} \log f_j\}$. This equation is correct subject to a fixed normalisation of the model $f_j = \sum_i \{\alpha_{ij} p(g|v_i)\}$ (in accordance with use of Extended Maximum Likelihood). We therefore perform normalisation on each model histogram so that the area under each model becomes equal to the number of corresponding data points. As this expression is proportional to the joint probability, the optimisation of this function is valid for parameter estimation. However, the unknown scale factor makes the measure unsuitable as an absolute estimate of fit quality (see below). The total cost function when summed over all image regions is $M_v = \sum_j \{L_j\}$. This expression is optimised using the downhill simplex method of Nealder and Meade [4], with restarts in order to avoid local minima.

Test phase: Once an approximate model is obtained, the optimisation process does not need to be executed again for the test data and estimated model parameters can be stored in a database. Then, for each new test image, we build J data histograms with specifications similar to those used in the training phase. Since grey levels stored in image files from different imaging equipment may have different scales, we apply a scale factor that is varied in the range [0.5, 2.0] to find the best fit of the input data to the model. Obviously, using the model histogram specifications some scales may result in overflow or underflow in the data histograms. These cannot correspond to the best fit and are ignored. A 10% tolerance on the model histogram range is used during the training phase. To obtain an absolute measure of similarity, the out-of-fit measure is then computed using the Matusita measure [7, 8] so that

$$M_v = (1/4JH) \sum_{j,g} \{[\sum_i \alpha_{ij} p(g|v_i)]^{1/2} - (f_{gj})^{1/2}\}^2 \tag{3}$$

where H is the number of bins for each histogram. This can be considered as a χ^2 test, (i.e. the $\sqrt{f_{gj}}$ values will closely approximate a Gaussian distribution with a σ of $1/2$).

As the search for the best corresponding scale is an optimisation with one parameter it is amenable to direct search. We set the scale step to 0.02 and compute the out-of-fit measure at 76 scales in the range [0.5, 2.0] (this involves no more evaluations than would be expected if using a conventional optimisation). One may proceed further by interpolating the minima from a quadratic equation to three points for increased accuracy.

3 Experiments

The aim is to gain parameter stability by obtaining multiple linearly independent examples of image histograms [2]. When sub-dividing an image into regions there is clearly a possible trade off between the number of regions and the resulting number of samples in each. We set the number of bins to 108 and divide each image into 4 by 4 windows which makes 16 corresponding histograms. We trained the model using a single slice from a 3D MR image of the normal human brain, shown in Fig. 2 (also shown larger in Fig. 1). The algorithm was converged with an out-of-fit measure at 0.6172.

To study how the out-of-fit measure behaves, we have also set the number of windows at 4, 6, 9, 12, 16, 20, 25 and 100. As expected, the larger the number of histograms the smaller the out-of-fit measure, and so more accurate fits are obtained. Of course, increasing the number of histograms to some extent is advantageous but having too many histograms lowers the ability of the model to differentiate between valid and invalid test images.

Valid test data: We tested 9 MR images against the model (see Fig. 2). The results are listed in Table 1. It is clear from this experiment that the out-of-fit measure in all cases is close to its value for training data. The deviation from the typical measure value is small for the whole set. One may also investigate training using several different images and using an average model. We perform further tests below using different data to evaluate the algorithm.

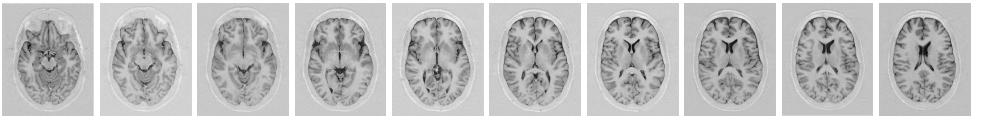

Figure 2: Valid MR brain image slices with results given in Table 1; slice numbers from left-to-right: 10, 11, 12, ..., 18 and 19; the model was trained using slice 12.

slice	10	11	12	13	14	15	16	17	18	19
scale	1.14	1.10	1.12	1.24	1.18	1.16	1.20	1.20	1.20	1.22
measure	0.67	0.56	0.51	0.51	0.53	0.55	0.55	0.60	0.68	0.87

Table 1: Test results on original data (trained using slice 12): rows refer to the image slice number, scale factor giving the best fit, and the corresponding out-of-fit measure.

Re-scaled test data: One issue of data quality that frequently occurs is that data is under-quantised during acquisition or following an image conversion for file storage. This often has negative effects on sophisticated analysis processes, particularly those that involve data density modelling or require spatial derivatives. Such a process directly modifies the structure of the image histogram and should be detectable via our quality checking process. A second experiment was performed in which the images from Fig. 2 were quantised at 32 grey levels, producing gaps in the histograms. Results are shown in Table 2. In comparison to table 1, the out-of-fit measure is significantly higher, confirming the ability of the proposed technique to detect this type of artefact.

Invalid test data: To test using some MR images of different imaging parameters or different tissues, we processed a variety of MR images so that their histograms ranges correspond to the range used during the training (Fig. 3). The results are shown in Table 3. Again, the out-of-fit measures are significantly higher than those found in Table 1, confirming the ability of the technique to detect application of an algorithm to invalid image type.

4 Conclusions

We have identified the problem of use of algorithms on data that is not suitable for such processing when analysis software is used as a measurement tool. Conventional approaches to the issue of quality control involve checking imaging parameters or signal to noise. Such tests are unlikely to identify more subtle problems, particularly when obtaining data from alternative imaging equipment. Unfortunately, such problems are often difficult to identify without significant technical knowledge and access to appropriate investigative tools. In order to deal with this problem we have suggested a supplementary statistical test based upon the construction of a component model, trained on sub-regions of images known to be suitable for analysis. We have shown how this technique will identify not only quantisation effects, but also novel histogram structure arising from different biological structures. [1]

slice	10	11	12	13	14	15	16	17	18	19
scale	1.005	1.01	1.008	1.01	1.007	1.01	1.01	1.011	1.005	1.24
measure	1.42	1.33	1.31	1.36	1.35	1.38	1.42	1.47	1.57	1.79

Table 2: Test results on re-scaled data (trained using slice 12): rows as Table 1.

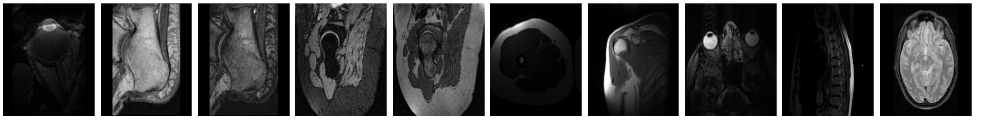

Figure 3: Invalid MR images (coil) with results given in Table 3; from left-to-right: eye, foot0, foot1, hip1, hip2, hip3, shoulder, skin, spine and brain-pd.

image	eye	foot0	foot1	hip1	hip2	leg	shoulder	skin	spine	brain
scale	0.52	0.76	0.54	0.55	0.52	0.52	0.59	0.54	0.52	0.68
measure	12.5	6.4	8.3	9.6	6.9	13.8	11.1	11.6	13.3	9.9

Table 3: Test results on re-scaled invalid data (trained using slice 12): rows as Table 1.

References

[1] R J Barlow. *Statistics: A Guide to the use of Statistical Methods in the Physical Sciences*. John Wiley and Sons Ltd., UK, 1989.

[2] P A Bromiley and N A Thacker. When less is more: Improvements in medical image segmentation through spatial sub-sampling. Technical Report TINA Memo no. 2007-005, The University of Manchester, 2007. `www.tina-vision.net/docs/memos/2007-005.pdf`.

[3] P A Bromiley and N A Thacker. Multi-dimensional medical image segmentation with partial volume and gradient modelling. *Annals of the BMVA*, 2008(2):1–22, 2008.

[4] W H Press, B P Flannery, S A Teukolsky, and W T Vetterling. *Numerical Recipes in C*. Cambridge University Press, New York, 2nd edition, 1992.

[5] K Rank, M Lendl, and R Unbehauen. Estimation of image noise variance. *IEE Proc Vis Image Signal Process*, 146(2):80–84, 1999.

[6] P Santago and H D Gage. Quantification of MR brain images by mixture density and partial volume modelling. *IEEE Trans Med Imaging*, 12:566–574, 1993.

[7] N A Thacker and P A Bromiley. The effects of a square root transform on a Poisson distributed quantity. Technical Report TINA Memo no. 2001-010, The University of Manchester, 2001.

[8] N A Thacker, F Ahearne, and P I Rockett. The Bhattacharryya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 34(4):363–368, 1997.

[1] This work was performed in collaboration with the Max Planck Institute for Evolutionary Biology, Germany.