CLASSIFICATION OF BLADDER CANCER ON RADIOTHERAPY PLANNING CT IMAGES USING TEXTURAL FEATURES

Hanqing Liao[1], William H. Nailon[1], Duncan B. McLaren[2] and Steve McLaughlin[1]

[1]Institute of Digital Communication, University of Edinburgh
AGB, King's Buildings, EH8 9UL, Edinburgh, UK
phone: + (44) 131 650 5659, email: h.liao@ed.ac.uk
[2]Directorate of Clinical Oncology, Western General Hospital
Crewe Road South, EH4 2XU, Edinburgh, UK phone: + (44) 1315373560, fax: + (44) 1315371092

ABSTRACT

Highly reliable classification of anatomical regions is an important step in the delineation of the gross tumour volume (GTV) in computed tomography (CT) images during radiotherapy planning. In this study pixel-based statistics such as mean and variance were insufficient for classifying the bladder, rectum and a control region. Statistical texture analysis were used to extract features from gray-tone spatial dependence matrices (GTSDM). The features were de-correlated and reduced using principal component analysis (PCA), and the principal components (PC) were classified by a naive Bayes classifier (NBC). The results suggests that the three most significant PC of the 56 features from GTSDM with distances $d = 1, 2, 3, 4$ give the highest average correct classification percentage.

1. INTRODUCTION

Accurate delineation of the GTV on CT images is vital in cancer radiotherapy planning to limit the radiation damage to normal tissue to maximize the dose to cancerous tissue. The International Commission on Radiation Units and measurements (ICRU)[1] defines the GTV as "the gross palpable or visible/demonstrable extent and location of the malignant growth." This is based on "purely anatomic-topographic and biological considerations without regard to technical factors of treatment." In the treatment of cancer by radiotherapy, CT images are used for treatment planning because they offer significant advantages over other imaging modalities such as magnetic resonance imaging (MRI). Firstly, CT images show superior consistent geometry, that is, they have less spatial distortion, consequently the volume obtained from CT images is more accurate, which is crucial for radiotherapy planning. Secondly, electron density information can easily be derived from CT images for accurate dose calculation. Thirdly, bones appears bright and in the CT images, which is important for identifying rigid landmarks and verifying setup accuracy[20].

However, the soft tissue contrast in CT images is relatively poor compared to MR. Determination of the GTV in CT images thus demands significant clinical-experience, and is extremely time-consuming, which leads to numerous problems. Firstly, significant inter- and intra-clinical variability of GTV has been reported in literature[21]. Secondly, because treatment periods are long, many factors such as patient movement may change the position of the GTV, resulting in less than optimal treatment. Furthermore, the widespread introduction of multi-slice CT places significant pressure on clinicians because the number of images that require outlining increases significantly. In light of these issues, there is a need for a reliable, objective method to assist clinicians contouring on CT images. Several methods, such as region growing, thresholding, Markov random field (MRF) models, classifying etc. have been proposed in the literature to find the GTV automatically, and Pham et al.[15] present a review of medical image segmentation methods.

Recently texture analysis methods have been reported as offering good classification performance, which will be reviewed in Section 2. This paper extends the work by Nailon et al.[14] by investigating different feature reduction and classification strategies and modifications to the texture analysis algorithms. Texture analysis is a set of computer image processing methods aimed at extracting the information required to represent textures as textural features. In this paper, texture analysis is used to find textural features that are similar among anatomical regions with similar pathology and distinct between different anatomical regions. Figure 1(a) shows a typical CT image from a bladder cancer patient. There are three regions of interest (ROI): bladder, rectum and a control region containing multiple pathology. For radiotherapy planning, the focus is on the delivery of as much dose to the tumour as possible while limiting the dose to surrounding organs. It would therefore be desirable to classify the GTV automatically with high degree of accuracy and reliability. To this end three distinct ROI from the bladder, rectum and a control region containing multiple pathology were investigated.

Two advances to the previous approach have been made. In the first, 56 features from four co-occurrence matrices with distance $d = 1, 2, 3, 4$ were used. In addition, the PCA was used to de-correlate the obtained features. The j most significant PC are then classified by NBC. The results show that the three most significant principal features can offer high correct rate classification of bladder, rectum and control region. Secondly the cross-validation experiments are conducted so that N different images are used as training set while thre rest images are used as testing set. The average correct classification percentage are high, which suggests the method is reliable. This approach has the potential to be used as part of an algorithm for assisting clinicians delineate the GTV.

The paper is organized as follows. In Section 2 the co-occurrence matrices texture analysis method is reviewed. Section 3-5 describes the methods of GTSDM, PCA and NBC. Section 6 illustrates the results and discusses the textures at different scales on CT images: the macro-textures

and micro-textures, and how they affect the GTSDM. Section 7 concludes the paper and discusses the future work.

2. LITERATURE REVIEW: GTSDM

Famous examples of texture images have been given by Brodatz[3], however there is no accurate definition on texture. The purpose of statistical texture analysis is to describe the characteristics of textured images by features, which can be used for classification. Haralick et al. proposed a method called gray-tone spatial dependence matrices (GTSDM), or co-occurrence matrices, to classify different textures[7]. This method characterizes texture by exploring the statistical properties of the spatial dependency of a pixel with its neighbours. Features from the GTSDM are reported to have high texture classification performance in comparative studies of different texture analysis methods[17, 19, 2]. Moreover, the GTSDM are also reported successful in classifying sonar [8] and radar[4, 18, 11, 5] images. The GTSDM method is the most heavily studied texture classification approach.

In medical image processing, Hamilton et al. [6] used features from the GTSDM approach to identify focal areas of colorectal dysplasia from a background of histologically normal tissue and reported an accuracy of 86% for the training data set and 83% for the large histological scene split into smaller component images. Koss et al. [12] applied the GTSDM method to an abdominal CT image to segment 7 different organs, and reported a successful percentage of 79 - 100%. Nailon et al. studied CT images of genitourinary cancer [14], and report features from GTSDM showed the best performance in classifying bladder and rectum regions. Philips et al. [16] form and examine 3-D liver CT images and reported a variation in accuracy from 84.663% to 89.459% by changing the directions of the GTSDM.

3. GTSDM CALCULATION AND FEATURE EXTRACTION

The texture analysis using GTSDM is used to characterize different ROI. According to [7], the co-occurrence $c_i j$ is defined as a function of gray-levels i, j of two pixels with distance d from each other in direction $\theta = 0°, 45°, 90°, 135°$. In this study it is found that there are no significant difference between GTSDM with different θ, so the four GTSDM with different θ were averaged for statistical consistency. In order to characterize the GTSDM, 14 statistical features defined in [7] were extracted from GTSDM with distance d. It is also found that different d offered additional information for classification, so in this experiment $d = 1, 2, 3, 4$ were used, and totally 56 features were extracted. In the following paragraphs, the features will be denoted as $\mathbf{x_m}(k)$, m is the total number of features for one ROI,.

4. FEATURE REDUCTION USING PCA

One problem with the statistical features defined in [7] is they may be correlated with each other. While auto-feature selection algorithms are proposed in the literatures[14], the de-correlation problem has not received much attention. In this study PCA is used to map the features into a linear subspace with minimum correlation in second-order sense. For the features $\mathbf{x_m}$ with covariance matrix Σ, the the kth principal component (PC) is given by $z_k = \alpha'_{\mathbf{k}} \mathbf{x_m}$, where $\alpha_{\mathbf{k}}$ is the eigenvector corresponding to the kth largest eigenvalue of Σ, and the variance of the PC z_k is the kth largest eigenvalue λ_k of Σ[10]. It can thus be inferred that PCA can maximally retain variation present in the dataset while reducing the number of features. In this study, the features were first normalized to the interval $[0, 1]$ to avoid pick up features large in number to be PC. The PC of $\mathbf{x_m}$ is denoted as $\mathbf{z_m}$, and the kth most significant PC is $\mathbf{z_m}(\mathbf{k})$.

5. NAIVE BAYES CLASSIFICATION

NBC is based on a simple assumption that the features are conditionally independent given the target classes C_i[13, 9]:

$$P(z_m(1), z_m(2)...z_m(n)|C_i) = \prod_{j=1}^{n} P(z_m(j)|C_i) \qquad (1)$$

The PCA can maximally de-correlate the features to meet the conditional independent assumption of NBC. In existing literature, Yu et al.[22] reported good performance using PCA-NBC jointly to classify aerial images. In this study, the PCA-NBC method is employed to evaluate the classification performance of statistical textural features from different ROI. In order to do this, the posterior probability $P(C_i|z_m(1), z_m(2)...z_m(n))$ is required. This can be calculated as follows:

$$
\begin{aligned}
p(C_i|z_m(1), z_m(2)...z_m(n)) &\propto p(C_i)p(z_m(1), z_m(2)...z_m(n)|C_i) \\
&\propto p(C_i)\prod_{j=1}^{n} p(z_m(j)|C_i) \\
&\propto \prod_{j=1}^{n} p(C_i|z_m(j))/p(C_i) \qquad (2)
\end{aligned}
$$

If for $j = 1...n$, $z_m(j) \in C_i$ is assumed to have a Gaussian distribution, the distributions in (2) can be inferred from a training set. New data can then be classified by maximum a posteriori (MAP) criterion.

6. RESULTS AND DISCUSSIONS

In this study 59 CT images acquired on 8 bladder cancer patients in different days during the treatment were examined. Images were scaled so that all pixels have positive gray levels. The centers of ROI in each image were given by experienced clinician, and for each ROI a 20-by-20-pixel area is examined. The means and variances of different ROI are shown in Figure 1(b). It can be seen that the means and variances of the three ROI have significant overlap, thus they cannot yield satisfactory classification result.

For 14 features from GTSDM with distance $d = 1$, Figure 2(a) illustrates the amplitude of the eigenvalues of Σ_{14}. It can be seen that there are five significant degrees of freedom in principal features, but according to Figure 2(b), the two most significant PC, $z_{14}(1)$ and $z_{14}(2)$, still cannot give a satisfactory classification result, since the distances between PC of different ROI are not large enough. Then 56 features from GTSDM with $d = 1, 2, 3 and 4$ were used for PCA, Figure 3(a) illustrates the amplitude of the eigenvalues of Σ_{56}, it is shown that the degrees of freedom increase, and the amplitude of variance in each PC subspace also increases. More-

over, Figure 3(b) gives a visualization of $z_{56}(2)$ and $z_{56}(3)$. it can be seen intuitively that the three ROI can be distinguished by the two PC.

The classification performance of different number of PC from GTSDM with $d = 1, 2, 3, 4$ were examined by using NBC. First, j most significant PC, i.e. $z_{56}(1)...z_{56}(j)$ were used for classification. Then the PC set with j variables and 59 records was randomly divided into the training group containing N records and the testing group with the rest records. By assuming all PC are conditionally independent and Gaussianly distributed, the prior probability $p(C_i)$ and the conditional distribution $p(C_i|z_{56}(k)), k = 1, ...j$ can be determined from the training set, and posterior probability $p(C_{bladder}|z_{56}(1), z_{56}(2)...z_{56}(j))$, $p(C_{rectum}|z_{56}(1), z_{56}(2)...z_{56}(j))$ and $p(C_{control}|z_{56}(1), z_{56}(2)...z_{56}(j))$ can thus be calculated for each testing record. Decisions can be made by MAP criterion. For each j-N setup, 500 cross validation experiments were conducted to evaluate the classification performance by using different training sets. The average correct classification rate are shown in Table 1. Significant classification performance can be achieved by using PC of statistical features from GTSDM with $d = 1, 2, 3, 4$. This further substantiates the assertion that the texture of different ROI contains important information for high-accuracy classification.

Table 1: Correct Classification Rate (%) using the j most significant PC and N Training Images

N	j = 1	j = 2	j = 3	j = 4	j = 5	j = 6
3	77	83	88	86	83	82
6	82	92	95	95	93	93
9	83	93	96	95	94	94
12	83	93	96	95	94	94
15	83	92	95	94	94	93
18	82	91	93	92	92	92

The highest classification rate occurred when $j = 3, N = 9$. By increasing the number of PC used for classification from 1 to 3, the correct percentage also increased. However, when more than 3 PC were used, the correct percentage dropped. According to the Bayesian probability rule, more variables should always increase the classification performance, but in most machine learning cases, it is impossible to know the underlying distribution of the variables, the assumed distribution will be invalid if too many irrelevant variables are used. For PCA, classification power will drop from $z_m(1)$ to $z_m(n)$ as the variances within the principal subspace decreases, therefore it is crucial to find the threshold j so that $z_m(1), ...z_m(j)$ offer the best classification performance.

7. CONCLUSIONS AND FUTURE WORK

In this study the importance of GTV auto classification has been addressed. One difficulty encountered is that the contrast of the soft tissue in CT images is poor, and the means and variances are insufficient for characterization of different ROI for classification. Texture analysis was used to classify different anatomical ROI on CT images. For 59 CT images from bladder cancer patients, high correct percentage classification is achieved by using NBC to classify three most significant PC of 56 features from GTSDM with $d = 1, 2, 3, 4$. As the GTSDM with different d describe spatial structure of the texture, the result proves that there is significant information in the texture of ROI for classification. In future work, the proposed method will be applied to the whole CT image to try to identify the ROI by features.

The proposed method follows the track of feature extraction - feature reduction process. However, the fact that only three PC were used suggests the GTSDM method requires more computational power than necessary. As PC represent the underlying latent variables, it is more promising to find the underlying latent variables directly to reduce the computational complexity. One possible approach is to define new features to characterize the GTSDM specially for ROI classification. This will also be examined in future work.

REFERENCES

[1] Prescribing, recording, and reporting photon beam therapy. Technical Report 50, Interational Commition on Radiation Units and Measurements, 7910 Woodmont Avenue, Bethesda, Maryland, 20814, USA, September 1993.

[2] J. Berry, J.R. and J. Goutsias. A comparative study of matrix measures for maximum likelihood texture classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 21(1):252–261, Jan/Feb 1991.

[3] P. Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover Publications Inc., 2000.

[4] L. Bruzzone, P. Pellegretti, and F. Roli. An experimental analysis of the use of grey level co-occurrence statistics for sar-image classification. *Geoscience and Remote Sensing Symposium, 1995. IGARSS '95. 'Quantitative Remote Sensing for Science and Applications', International*, 2:1431–1433 vol.2, 10-14 1995.

[5] J. Carr and F. de Miranda. The semivariogram in comparison to the co-occurrence matrix for classification of image texture. *Geoscience and Remote Sensing, IEEE Transactions on*, 36(6):1945–1952, Nov 1998.

[6] P. W. Hamilton, P. H. Bartels, D. Thompson, N. H. Anderson, R. Montironi, and J. M. Sloan. Automated location of dysplastic fields in colorectal histology using image texture analysis. *The Journal of Pathology*, 182(1):68–75, 1997.

[7] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 3(6):610–621, Nov. 1973.

[8] K. Imen, R. Fablet, J.-M. Boucher, and J.-M. Augustin. Statistical discrimination of seabed textures in sonar images using co-occurrence statistics. *Oceans 2005 - Europe*, 1:605–610 Vol. 1, June 2005.

[9] W. Q. A. M. S.-A. M. Islam, M.J. Investigating the performance of naive- bayes classifiers and k- nearest neighbor classifiers. In *Convergence Information Technology, 2007. International Conference on*, pages 1541 –1546, Nov. 2007.

[10] I.T.Jolliffe. *Principal Component Analysis*. Springer, second edition edition, 2002.
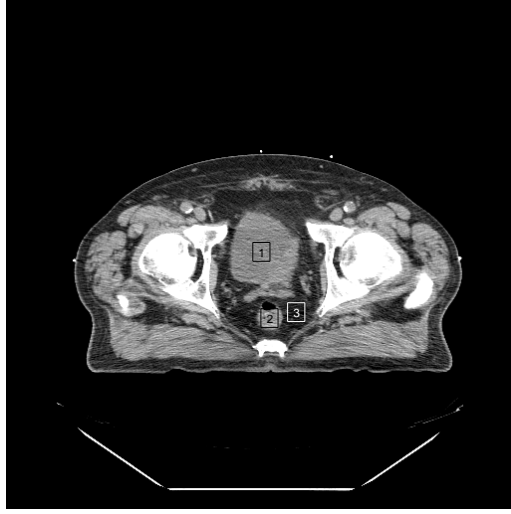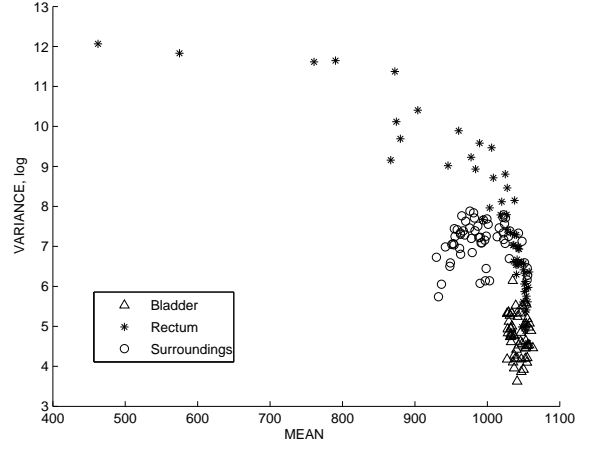
Figure 1: Left (a): A typical CT Image for Radiotherapy Planning containing three Regions of Interest (ROI): 1. Bladder, 2. Rectum and 3. a control region. Right (b): The Means and Variances (log) of the Three ROI from 59 radiotherapy planning CT images. Significant overlap can be observed, so Different ROI cannot be distinguished by Means and Variances.
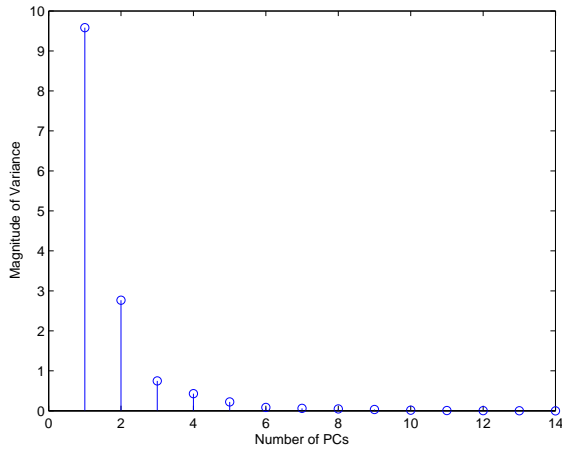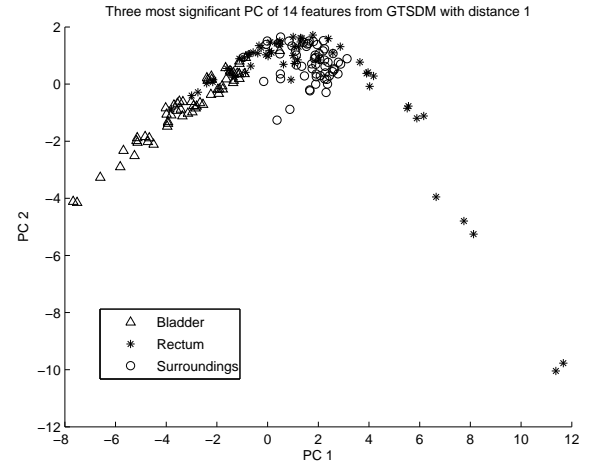
Figure 2: Left (a): The Eigenvalues of the Covariance Matrix Σ_{14} features from GTSDM with $d = 1$, Representing the Variance in each principal subspace. Five Significant Degree-of-Freedom can be observed. Right (b): Visualizing PC 1 against PC 2: ROI still cannot be distinguished intuitively.
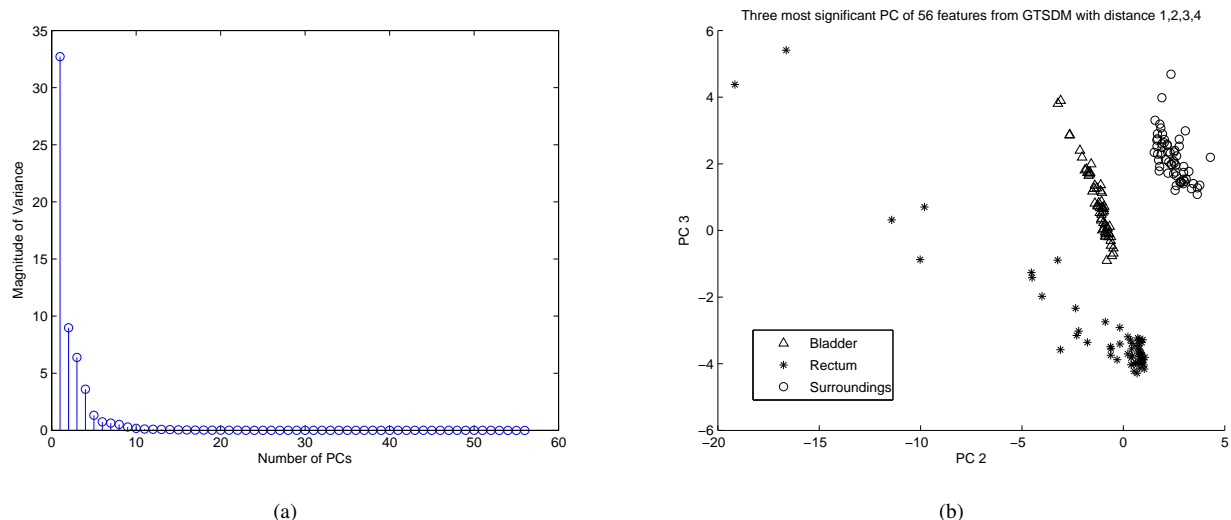
(a)

Three most significant PC of 56 features from GTSDM with distance 1,2,3,4

(b)

Figure 3: Left (a): The Eigenvalues of the Covariance Matrix $\mathbf{\Sigma_{56}}$ from GTSDM with $d = 1, 2, 3, 4$, Representing the Variances in each principal subspace. Six Significant Degree-of-Freedom can be observed, Comparing with Figure 2(a), the Amplitude of Variance in Each PC subspace increases significantly. Right (b): Visualizing PC 2 against PC 3: ROI can be distinguished intuitively. Further Classification Performances using Different Number of PC are Described in Table 1.

[11] U. Kandaswamy, D. Adjeroh, and M. Lee. Efficient texture analysis of sar imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(9):2075–2083, Sept. 2005.

[12] J. Koss, F. Newman, T. Johnson, and D. Kirch. Abdominal organ segmentation using texture transforms and a hopfield neural network. *Medical Imaging, IEEE Transactions on*, 18(7):640–648, July 1999.

[13] S. L. Martinez-Arroyo, M. Learning an optimal naive bayes classifier. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pages 958 –958, 0-0 2006.

[14] W. Nailon, A. Redpath, and D. McLaren. Texture analysis of 3d bladder cancer ct images for improving radiotherapy planning. *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 652–655, May 2008.

[15] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. In *Annual Review of Biomedical Engineering*, volume 2, pages 315–338. 2000.

[16] C. Philips, D. Li, D. Raicu, and J. Furst. Directional invariance of co-occurrence matrices within the liver. *Biocomputation, Bioinformatics, and Biomedical Technologies, 2008. BIOTECHNO '08. International Conference on*, pages 29–34, 29 2008-July 5 2008.

[17] C. H. Richard Conners. A theoretical comparison of texture algorithm. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol.PAMI-2(3), May 1980.

[18] L.-K. Soh and C. Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(2):780–795, Mar 1999.

[19] J. S. Weszka and A. Rosenfeld. A comparative study of texture measures for terrain classification. *NASA STI/Recon Technical Report N*, 76:13470–+, Mar. 1975.

[20] R. Williams, J and I. Thwaites, D, editors. *Radiotherapy Physics: in Practice*. Oxford University Press, second edition edition, 2000.

[21] M. Yamamoto, Y. Nagata, K. Okajima, T. Ishigaki, R. Murata, T. Mizowaki, M. Kokubo, and M. Hiraoka. Differences in target outline delineation from ct scans of brain tumours using different methods and different observers. *Radiotherapy and Oncology*, 50(2):151 – 156, 1999.

[22] X. Yu, Z. Zheng, L. Li, and Z. Ye. Texture classification of aerial image based on pca-nbc. In L. Zhang, J. Zhang, and M. Liao, editors, *Proceedings of SPIE, the International Society for Optical Engineering*, volume 6043, page 60432G. SPIE, 2005.