Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Content Based Image Retrieval for MR Image Studies of Brain Tumors

Shishir Dube[1], Suzie El-Saden[1], Timothy F. Cloughesy[2], Usha Sinha[1]

[1]Medical Imaging Informatics Group (MII)
[2]Department of Neurology
University of California at Los Angeles (UCLA)

Abstract- This work proposes a methodology for content-based image retrieval of glioblastoma multiforme (GBM) and non-GBM tumors. Regions containing GBM lesions from 40 patients and non-GBM lesions from 20 patients were manually segmented from MR imaging studies (T1 post-contrast and T2 weighted channels) to form the training set. In addition to the two acquired channels, a composite image was formed by an image fusion method. Data reduction techniques, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), were applied on the training sets (T1 post, T2, composite, and multi-channel combining the PCA features from T1post and T2). The retrieval accuracy was evaluated using a 'leave-one-out' strategy with query images belonging to 'normal', 'GBM' and 'non-GBM' classes. Several combinations of the similarity metric and classifier were used: Euclidean similarity measures with k-means classifier for the PCA and LDA features and Support Vector Machine (SVM) nonlinear classifier (radial basis function kernel) with the PCA derived features. The SVM classifier served as a comparison of nonlinear techniques vs. linear ones. Multi-channel PCA was 100% accurate in classifying a query image as either 'normal' or 'abnormal'. The highest accuracy in classification of tumor grade (GBM or other Grade 3) was 77% and was achieved by SVM coupled with the PCA features. The proposed algorithm intent is to be integrated into an automated decision support system for MR brain tumor studies.

I. Introduction

Medical image databases have grown immensely in the past few years. Imaging studies, such as magnetic resonance (MR) and computed tomography (CT), result in a large volume of data. The ability to query large image databases can form the basis of a decision support system. However, a large number of existing image databases are indexed by text annotations that routinely contain only patient demographic or study information (e.g., age, gender, date of study, modality). Manual methods to extend this with more comprehensive text annotation that capture image content are both tedious and time-consuming [1]. On the other hand, Content Based Image Retrieval (CBIR) systems allow users to query based on the image content (i.e. image-derived features) rather than the associated text annotation. Research in effective medical CBIR systems is an active and ongoing area for the past decade and several systems are currently being evaluated for clinical retrieval and classification tasks: ASSERT, IRMA, and medGIFT.

A CBIR system for brain lesions has been proposed earlier based on the analysis of texture and histogram features derived from subtracted half brains. The system was able to classify 2D CT slices as normal, tumor, hemorrhage, or stroke. We propose a CBIR system for brain tumors from multi-channel MR images: the system can classify a user-selected region as normal or abnormal; and if abnormal, further classify as GBM or non-GBM. The advantages of the system that we propose mainly lie in diagnosing between grades of tumor with minimal invasive procedures. The algorithm also retrieves from the training set an image with the closest visual match to the query region. The algorithm is based on data reduction techniques: principal component analysis (PCA) and linear discriminant analysis (LDA) performed directly on the image region intensities. The algorithm was tested on each channel (T1 post, T2 images) independently, as well as the combined features of both channels (termed multi-channel) and finally on the composite image generated from the two channels [2]-[4].

II. Methodology

Clinical MR brain images of subjects with GBM and other Grade 3 tumors (Anaplastic Astocytoma, Anaplastic Mixed Glioma, and Anaplastic Oligodendroglioma) were selected from the Brain Tumor Database at UCLA. The diagnosis of all subjects was confirmed with histopathology. The axial image studies included in this report were limited to pre-surgery first presentation cases. This circumvented the problem of tumor alteration due to surgery and/or chemotherapy and radiation. Several pre-processing steps (detailed below) were performed prior to creation of the training sets to account for sub-optimal image quality, non-alignment of images within a study and varying image intensities between studies. Regions of normal and abnormal tissue were identified manually from these processed images to create separate training sets of independent channels and the fused images. The training set consisted of 40 pre-operative patient imaging studies diagnosed with GBM and 20 pre-operative patient imaging studies diagnosed with non-GBM tumors.

Preprocessing:

The first step in the methodology was pre-processing the training data that would become a part of the brain tumor database in our CBIR framework. All image data sets consisted of two contrasts (or channels): T1 post contrast enhanced and T2 weighted images. The following steps were performed to improve SNR, remove non-brain voxels, register the two image channels of a given subject, and to obtain consistent image intensities across all subjects for a given channel:

- Denoising –FSL's (**F**MRI **S**oftware **L**ibrary) nonlinear filter which is based on averaging voxels with local voxels that have similar intensity values [5]
- Brain Extraction – Removal of all non-brain matter with the BET algorithm from the FSL package [5]
- Registration – The non-contrast enhancing brain volumes were linearly aligned with an affine transformation to the respective T1 post contrast brain volume (within the same patient) via FSL since brain volumes within the same patient are not always aligned [5]
- Intensity Standardization – Performed via a histogram matching method. Images from each study acquired at a given contrast were matched to pre-determined histogram (providing a tissue specific meaning) for that contrast. The pre-determined histogram was an average of several matched histograms of 10 randomly chosen subjects [6].

Composite Images:

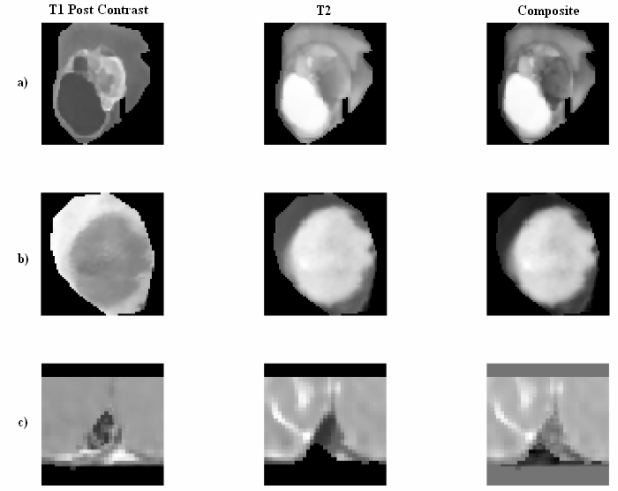

Figure 1: a) GBM b) Grade III and c) Normal tissue (located in the posterior part of the brain) examples of images in three presentations (T1 post contrast, T2, and composite labels along the columns).

Figure 1 visualizes how an image fusion method may provide a robust way of combining information from several channels into a compact scene. For the GBM example (row a), one can see that the composite image (third column) has a clearer border between the tumor and edema segments. The non-GBM abnormal class (row b) image has a composite image, which highlights only the lesion on the T2 channel with no surrounding edema. Row c is the normal tissue class.

We formed a composite image from the two channel data that would capture the salient contrast features of all channels into one image. PCA was applied to form composite images of the multi-channel segmentations in the following manner:

- Form a matrix X_{NP} to describe the data set, where each row of the matrix gives intensity values of one of the N voxels for each of the P images in the multi-channel set to satisfy the following: $Y_{N1} = X_{NP}W_{P1}$, where Y_{N1} is the single composite image
- Mean-correct and normalize the data matrix X_{NP} to prevent voxels of significantly high values of particular channels having extra emphasis
- Find the eigenvectors of the covariance matrix $X_{NP}{}^{T}X_{NP}$ such that: $X_{NP}{}^{T}X_{NP}e_i = \lambda_i e_i$ for $i = 1, \ldots, P$
- The eigenvector corresponding to the highest eigenvalue is set as the weighting vector W_{P1} to be applied to the multi-channel data to form the composite image

Since the maximum total voxel variance is equal to the largest eigenvalue, the proposed fusion scheme will result in a composite image with the highest contrast to noise ratio possible with a linear combination of the multi-channel images. Composite images were used to create training sets; PCA and LDA data reduction methods were applied to the training sets of composite images [7].

Creating the basis sets from the training set images using PCA and LDA:

After all imaging studies were pre-processed, abnormal regions (which included the contrast-enhancing tumor, necrotic area, and edema in the same image were manually segmented). The set of regions for each of the two channels formed two training sets. It should be noted that a region had to be segmented on only one channel since the other channels are aligned through the pre-processing step. The normal sub-regions that were manually segmented consisted of healthy gray matter/white matter, enhancing blood vessel regions, cerebrospinal fluid areas, and eye areas. All normal sub-regions were also size normalized to dimensions of 64x64. The scaling information for resizing was preserved so that size information could be retained. However, since the current training set is small, we retrieve closest matches based only on visual similarity independent of size.

III. Results and Discussion

A combination of 6 training sets were created from a total of 140 GBM tumor regions, 140 non-GBM tumor regions, and 100 normal regions for each of the two channels as well as from the composite images. Three training sets (T1 post contrast, T2, and composite) included all normal and abnormal images. This set was used to classify a query region as normal/abnormal. The second group of three training sets (T1 post contrast, T2, and composite) included only the abnormal regions. The second group was used to classify an abnormal region as Grade 4 (GBM) or other Grade 3 (Anaplastic Astrocytomas, Anaplastic Mixed Gliomas, and Anaplastic Oligodendrogliomas). The system was evaluated using a leave-one-out strategy with an entire patient study's images excluded from the training set when it was used as a query image.

Dimension Reduction	Channel	Classifier	Recognition Rate
PCA	T1 Post Contrast	Nearest Neighbor (Euclidean)	0.94
PCA	T2	Nearest Neighbor (Euclidean)	0.99
PCA	Composite	Nearest Neighbor (Euclidean)	0.98
PCA	Multi-channel	Nearest Neighbor (Euclidean)	1.00
PCA	T1 Post Contrast	LDA	0.47
PCA	T2	LDA	0.48
PCA	Composite	LDA	0.56
PCA	Multi-channel	LDA	0.46
PCA	T1 Post Contrast	SVM	0.75
PCA	T2	SVM	0.75
PCA	Composite	SVM	0.75
PCA	Multi-channel	SVM	0.75

Table 1: Table of classification accuracy of a set of classifiers for a training set consisting of images labeled as normal or abnormal. Training set size was 366 images.

Dimension Reduction	Channel	Classifier	Recognition Rate
PCA	T1 Post Contrast	Nearest Neighbor (Euclidean)	0.69
PCA	T2	Nearest Neighbor (Euclidean)	0.52
PCA	Composite	Nearest Neighbor (Euclidean)	0.52
PCA	Multi-channel	Nearest Neighbor (Euclidean)	0.58
PCA	T1 Post Contrast	LDA	0.54
PCA	T2	LDA	0.46
PCA	Composite	LDA	0.59
PCA	Multi-channel	LDA	0.45
PCA	T1 Post Contrast	SVM	0.77
PCA	T2	SVM	0.37
PCA	Composite	SVM	0.50
PCA	Multi-channel	SVM	0.53

Table 2: Table of classification accuracy of a set of classifiers for a training set consisting of only grade 3 and grade 4 tumors. Training set size was 273 images.

Table 1 lists the accuracy of classifying a query sub-region as normal or abnormal. In all, 100 normal images and 280 abnormal images were used as query images in this part of the validation. The multi-channel Euclidean distance similarity measure on the PCA features had the best performance at 100% accuracy for classifying an image as normal or abnormal. It is surprising since the LDA and SVM methods are better at discrimination/classification. However, in face recognition tasks, when the images (used for query and training) are well conditioned (as accomplished by the post-processing methods), PCA methods have been shown to outperform when the training set is small [8]-[10]. All channels showed nearly the same performance and this is not surprising since normal/abnormal regions can usually be discriminated equally well in all channels. However, a small advantage of the T2 channel can be seen in these images. The composite and multi-channel seem to have a small edge over the single channel images for the LDA and PCA methods respectively.

Table 2 shows classification for abnormal regions as Grade 3 or Grade 4. For this validation, only the training set images of abnormal images were used and in all 280 query images were used to assess the performance. The best classification between grade 3 and grade 4 tumors was obtained with SVM applied on the T1 post contrast PCA features with a recognition rate of 77%. The second best classification was an application of a Euclidean nearest neighbor algorithm applied on the PCA features of the T1 post contrast channel (which yielded a recognition rate of 69%). The significant improvement in classification via SVM shows that the discrimination between different lesion types (from just image intensities) is not accurate via linearly separable methods (such as LDA). A noteworthy observation was that a majority of the misclassifications occurred when the GBM queries retrieved AA class images, and AA queries retrieved GBM class images (which can be more easily visualized in Figure 2).

The reason for this confounding classification is apparent by an inspection of Fig.2, which highlights the difficulty in distinguishing between GBM and AA lesions. Several aspects of the clusters in the PCA space can be related to the appearance of different tumors: although GBM tumors exhibit common characteristics such as necrosis, ring enhancement, and peripheral edema, they still have a highly variable appearance. Thus, they do not cluster as tightly as the AMG and AO class samples. Some tumors from the AA class exhibit intensity patterns that are similar to GBMs and can be seen within the GBM clusters. The appearance similarity of these two tumor classes is not surprising since AA type tumors are known to progress into GBM, and exhibit characteristics such as contrast enhancement. It is likely that extraction of detailed features of each sub-region of the tumor may result in more accurate discrimination between the GBM and AA classes. The best channel for discrimination was the T1 weighted channel since, in this image type, the contrast enhancing part, which is the distinctive feature of GBM, is highlighted. The T2 channel, on the other hand, shows a hyper-intensity for both the edema and tumor with only subtle distinctions in intensity between the tumor sub regions. Thus the T2 channel is not good for discriminating the different tumor classes and may even contribute to noise as seen in the decreased performance of the multi-channel data (Table 2).

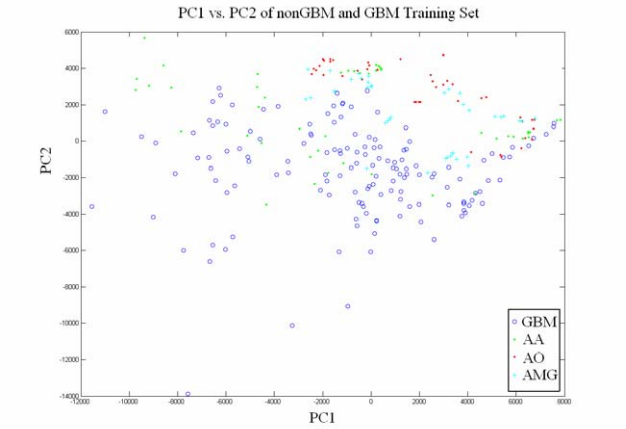

Figure 2: Plot of principal component 1 vs. principal component 2 of the T1 post contrast channel. The training set consisted of only abnormal tissues.

This algorithm is intended to be a part of a CBIR system that can evaluate a new MR imaging study of the brain for abnormality. In the current implementation, a user has to delineate a region of interest. However, we are also implementing a fully automated system where a new imaging study will be roughly clustered into four classes and each class will be first evaluated as normal/abnormal before proceeding to the second stage of classification as Grade 3 or Grade 4 tumor.

Further work will involve the use of other non-linear classifiers such as kernel PCA that have shown much promise in classification. Future work will also involve expanding the training sets for both GBM and non-GBM tumors, applying texture and shape feature filters for higher classification accuracy of tumors, and wavelet based image fusion techniques as an additional visual component for identification of tumors in a single scene.

REFERENCES

[1] Sinha, U, Kangarloo H. Principal Component Analysis for Content-based Image Retrieval, Radiographics, 22(5): 1272-1289,2002.

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval: The End of the Early Years, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12): 1349-1380, December 2000.

[3] Shyu, C, et al. ASSERT: A physician-in-the-loop content-based image retrieval system for HRCT image databases, *Computer Vision and Image Understanding*, 75(1/2): 111-132, July 1999.

[4] Dahmen J, Theiner T, Keysers D, Ney H, Lehmann TM, Wein BB. Classification of radiographs in the image retrieval in medical application system IRMA, Proceedings 6th International RIAO Conference on Content-Based Multimedia Information Access, Paris, France, 551-566, 2000.

[5] S.M. Smith, et al. Advances in functional and structural MR image analysis and implementation as FSL, NeuroImage, 23(S1): 208-221, 2004.

[6] Nyul LG, Udupa JK. On Standardizing the MR image intensity scale, Magn Reson Med., 42(6): 1072-81, 1999.

[7] Zhang Y, Goldszal A, Butman J, and Choyke P. Improving Image Contrast Using Principal Component Analysis for Subsequent Image Segmentation, Journal of Computer Assisted Tomography, 25 (5): 817-822, 2001.

[8] Turk M and Pentland A. Eigenfaces for Recognition, J. Cognitive Neuroscience, 3(1), 1991.

[9] Belhumeur P, Hespanha J, and Kriegman D. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19 (7): 711-720, 1997.

[10] Martinez A and Kak A. PCA versus LDA, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (2): 228-233, 2001.