BRAIN TUMOR CLASSIFICATION USING SPARSE CODING AND DICTIONARY LEARNING

Saif Dawood Salman Al-Shaikhli, Michael Ying Yang, Bodo Rosenhahn

Institute for Information Processing, Leibniz University Hannover
Appelstr. 9A, 30167 Hannover, Germany

ABSTRACT

Brain tumor classification is considered as one of the most challenging tasks in medical imaging. In this paper, a novel approach for multi-class brain tumor classification based on sparse coding and dictionary learning is proposed. We propose an individual (per-class) dictionary learning and sparse coding classification using K-SVD algorithm. This approach combines topological and texture features to build and learn a dictionary. Experimental results demonstrate that the sparse coding based classification outperforms other state-of-the-art methods.

Index Terms— Brain tumor classification, dictionary learning, sparse coding, topological matrix, gray level co-occurance matrix

1. INTRODUCTION

Early identification of brain tumors is important to treat the tumors effectively. Multi-class brain tumor classification is considered as one of the most important and challenging tasks in medical imaging due to the difficulty to extract the relevant information that can help to discriminate the tumor from the normal brain tissue [2]. Brain tumor classification involves two steps, feature extraction and classification. Feature extraction is an essential step in the classification since the relevant information from the original image needs to be chosen in order to achieve high brain tumor classification accuracy [1]. In general, brain tumors have different shapes and intensities from patient to patient [2], and sometimes, they also have different gray scales yet the same intensities as brain tissues [2]. Therefore, features related to the shape or intensity create ambiguities during tumor classification [2].

Thiagarajan et al. [4] proposed a sparse coding for brain tumor segmentation using intensity and location features. Bauer et al. [5] developed a fully automatic algorithm for brain tumor segmentation and classification using a support vector machine (SVM) with a hierarchical conditional random field. Han et al. [6] proposed an algorithm for gliobastoma multiforme classification in the histological images based on dictionary learning and sparse coding. The sparse coding based classification was compared with the traditional kernel methods of classification. They concluded that the accuracy of kernel methods are better than sparse coding for histological images. Selvaraj et al. [7] proposed an automatic classification technique based on Least Squares SVM to identify normal and abnormal slices of brain MRI images. Moon et al. [8] proposed an automatic brain tumor segmentation based on statistical classification with a geometrical prior. Cocosco et al. [9] proposed a fully automatic generation of correct training samples for MRI tissue classification. Weiss et al. [10] proposed an approach for multiple sclerosis lesion segmentation using dictionary learning and sparse coding using intensity features. In the previous papers, these approaches used either intensity-based or texture-based feature extraction for brain tumor classification, however a brain tumor may have the same intensity as normal brain tissue [11]. Furthermore, sparse representation has been shown to be an effective method for brain tumor classification by representing the images as dictionaries consist of linear combination of a few columns (atoms) of some redundant basis [16]. While in the Linear-SVM, the data may not be linearly separable in the original feature space and needs higher dimensional space mapping to increase the classification accuracy which is computationally expensive [6].

In contrast to previous papers, our **contribution** is a modified sparse coding and dictionary learning based multi-class classification. We proposed to use the K-SVD method to update both of the dictionary and sparse coding steps. Furthermore, due to the high degree of similarity in pixel intensities between normal brain tissue and tumor, and the variability of the tumor shape, location, and size, this variability justifies the use of topological and texture features to learn the dictionary. The topological feature gives information whether the case is normal or abnormal based on the assumption that the topology of normal brain is fixed. Therefore, the presence of tumor in the brain will change the normal brain topology. In addition, the texture features provide a good discrimination of the brain tumor types. The main novelty in our algorithm is the use of topology and texture features for learning, instead of applying learning directly on pixel values.

The work was partially funded by DAAD scholarship (A/10/96106) and MOHESR-Iraq (Baghdad University). The authors gratefully acknowledge these supports.

The rest of this paper is organized as follows: Section 2 explains the proposed method. Section 3 discusses the results and Section 4 summarizes the paper.

2. METHOD

2.1. Feature Extraction

In this subsection, the feature extraction step is explained by proposing a set of topological and texture features that give relevant information about the tumor.

2.1.1. Topological Matrix (TM)

The proposed topological matrix is represented by a topological graph relationship and it considers the main feature to classify the normal and abnormal brain images by assuming that the topology of the normal brain is fixed. The topological graph is constructed from the input data to provide the feature knowledge to the classifier. To compute TM, we consider an image I as sets of clusters depending on the dissimilarity between them $I = O_i, O_{i+1}, ..., O_N$. These clusters are connected with each other by a specific topological relationship. The clusters in the topological graph of the image I are computed using *Otsu's* method [12] and the topological relationship of these clusters are computed using the method proposed by Al-Shaikhli et al. [13]. Let O° be the interior of the cluster, ∂O be the boundary of the cluster, and χ_{O_i} is the membership function of each cluster. The topological relationship between the clusters is calculated in terms of probability of intersections of these clusters [13]:

$$V_{TM}(O_i, O_{i+1}) = (m_{11}, m_{12}, m_{13}, \ldots, m_{33})^T \quad (1)$$

V_{TM} in Eq. (1) is a vector of zeros and ones and it is the sum of all individual V_{TM} that are computed for each region ($V_{TM} = \sum_{i=1}^{N} V_{TM_i}$). The elements (that have ones values) represent the topological relationship of each region in the image. In our calculation, we consider only four elements (m_{11}, m_{12}, m_{21}, and m_{22}) and the rest are set as ones:

$$\begin{cases} m_{11}=0, m_{12}=0, m_{21}=0, m_{22}=0 & \text{if } RL_{dis}(O_i, O_{i+1}) > 0 \\ m_{11}=0, m_{12}=0, m_{21}=0, m_{22}=1 & \text{if } RL_{con}(O_i, O_{i+1}) > 0 \\ m_{11}=0, m_{12}=0, m_{21}=1, m_{22}=0 & \text{if } RL_{in}(O_i, O_{i+1}) > 0 \\ m_{11}=1, m_{12}=1, m_{21}=1, m_{22}=1 & \text{if } RL_{ov}(O_i, O_{i+1}) > 0 \end{cases} \quad (2)$$

where RL_{dis}, RL_{con}, RL_{in}, and RL_{ov} are disjoint, contact, inside and overlap region relationship respectively as follows:

$$RL_{dis}(O_i, O_{i+1}) = 1 - \max_b\{|\chi_{O_i}(b) + \chi_{O_{i+1}}(b) - 1|\} \quad (3)$$

$$RL_{in}(O_i, O_{i+1}) = \min(1, \min_b(1 + \chi_{O^\circ_{i+1}}(b) - \chi_{O_i}(b))) \quad (4)$$

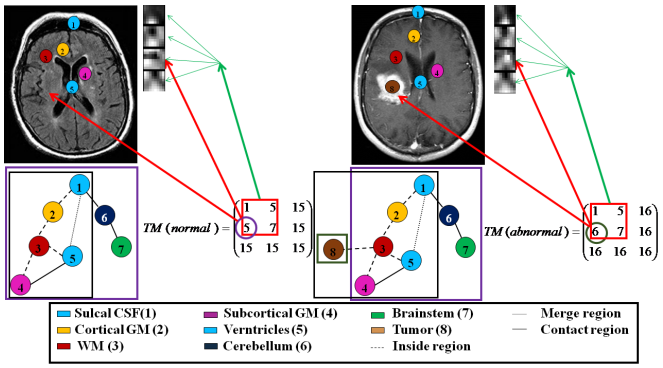

Fig. 1. Example of normal and abnormal brain MRI images with their topological graph and basis vectors of the topological feature. The element $m_{21} = 5$ in the normal case while $m_{21} = 6$ in the abnormal case, this indicates that there is a tumor in the brain. The violet rectangle represents the overall topological graph of the normal case, the black rectangle represents the topological graph of the example images, and the green rectangle represents the abnormal connectivity of WM.

$$RL_{con}(O_i, O_{i+1}) = \min\{(1 - \max_b(|\chi_{O_i^\circ}(b) + \chi_{O_{i+1}^\circ}(b) - 1|)), \max_b(\min(\chi_{\partial O_i}(b), \chi_{\partial O_{i+1}}(b)))\} \quad (5)$$

$$RL_{ov}(O_i, O_{i+1}) = \min\{\max_b(\min(\chi_{O_{i+1}^\circ}(b), \chi_{O_i^\circ}(b))), \max_b(\min(\chi_{O_i^\circ}(b), \chi_{\partial O_{i+1}}(b))), \max_b(\min(\chi_{O_{i+1}^\circ}(b), \chi_{\partial O_i}(b))), \max_b(\min(\chi_{\partial O_i}(b), \chi_{\partial O_{i+1}}(b)))\} \quad (6)$$

where b is a pixel in I. Table 1 and Fig. (1) illustrate the proposed topological properties for both normal and abnormal cases of the brain. In Table 1, the connected components represent the total relationship of each region. The number of cavities in each region indicates the number of regions inside it. In Fig. (1), the label (8) represents the abnormal connectivity of the white matter (presence of a tumor). Therefore the topological relationship of the white matter is changed. According to Eq. (1) and Eq. (2) this change is illustrated in element m_{21} in TM because the tumor is inside WM, for more details see [13]. This could be also seen in the basis vector of the topological feature of normal and abnormal brain MRI images.

2.1.2. Gray Level Co-occurance Matrix (GLM)

GLM is an important method for textural feature extraction proposed by Haralick et al [14]. Four texture features (contrast, correlation, energy, and inverse difference moment) are

Table 1. Topological properties for the normal (abnormal) cases.

Tissue label	Tissue type	# of connected components	Internal cavity	Handles
1	Sulcal CSF	3 (3)	1 (1)	1 (1)
2	Cortical gray matter (GM)	2 (>2)	1 (>1)	1 (>1)
3	White matter (WM)	3 (>3)	2 (>2)	2 (>2)
4	Subcortical gray matter	2 (>2)	0 (>0)	0 (>0)
5	Ventricles	3 (>3)	0 (0)	0 (0)
6	Cerebellum	2 (>2)	0 (>0)	0 (>0)
7	Brain stem	1 (>1)	0 (>0)	0 (>0)

considered for brain tumor classification. These features have been calculated for four different offsets ($0°$, $45°$, $90°$, and $135°$).

2.2. Dictionary Learning

In this subsection, the dictionary learning step in our algorithm using a K-SVD method will be presented to learn and update the dictionary. Let $c = 1, ..., 4$ is the number of the class, N_c are the training images of each class. D_c are the dictionaries of the corresponding training images of each class, and N is the sum of the training images of all four classes as explained in Fig. (2) which illustrates the proposed algorithm.

Let D_c be a dictionary $n \times K_c$ matrix $D_c = (d_1, d_2, ..., d_{K_c})$, which consists of K_c atoms (columns), $\{d_i \in R^n : i = 1, 2, ..., K_c\}$ and each atom represents the key features extracted from Y_c, where $(K_c \ll N_c)$ $Y_c = (y_1, y_2, ..., y_{N_c})$ is a $n \times N_c$ matrix which consists of feature vectors $\{y_i \in R^n : i = 1, 2, ..., N_c\}$ of N_c data samples (feature vectors) with dimension n. To compute the sparse representation $A_c = (a_1, a_2, ..., a_{N_c}) \in R^{K_c \times N_c}$, s.t. $y_i = D_c a_i$ and $\|a_i\|_0 << K_c, i = 1, ..., N_c$, the dictionary D_c by feature samples Y_c needs to be trained. In such a way that each feature vector in Y_c is represented by linear combination of a few atoms in the dictionary according to the non-zero elements in A_c as illustrated in the generative learning step in Fig. (2).

Our goal is to update the dictionary and the sparse representation A_c by minimizing the following equation using the K-SVD method [15]:

$$\arg \min_{D_c, A_c} \|Y_c - D_c A_c\|_F^2 \;\; s.t. \;\; \forall 1 \le i \le N_c, \;\; \|a_i\|_0 \ll K_c \tag{7}$$

To get an update of the dictionary D_c and the sparse representation A_c, we assume that the condition in Eq. (7) is a $K_c \times N_c$ matrix multiplied by A_c as a dot product of multiplication:

$$P_c = \begin{cases} P_c(i,j) = 1 & \text{for} \quad A_c(i,j) = 0 \\ P_c(i,j) = 0 & otherwise \end{cases} \tag{8}$$

Now, we can rewrite Eq. (7) as follows:

$$\{\hat{D}_c, \hat{A}_c\} = \arg \min_{D_c, A_c} \|Y_c - D_c A_c\|_F^2 \;\; s.t. \;\; P_c \circ A_c = 0 \tag{9}$$

The dot product $(P_c \circ A_c = 0)$ achieves all zeros in A_c without change. Equation (9) represents the update stage of the dictionary and we solve it by considering $D_c A_c$ as a sum of rank-1 outer products:

$$\{\hat{D}_c, \hat{A}_c\} = \arg \min_{D_c, A_c} \|Y_c - \sum_{i=1}^{K} d_i a_i^T\|_F^2$$
$$s.t. \;\; \forall 1 \le i \le K_c, \;\; p_i \circ a_i = 0 \tag{10}$$

To optimize the above equation, we use a block coordinate descent method. By multiplying the $(n \times N_c)$ rank-1 matrix $(1_n \cdot p_j^T)$ with Eq. (10), we compute the error matrices. Therefore, all columns of the samples that do not use j^th atom are removed.

$$E_i = (Y - \sum_{i \ne j} d_i a_i^T) \circ (1_n \cdot p_j^T) \tag{11}$$

where E_i are the overall representation error matrix. In Eq. (11), the rank-1 matrix represents the n times replication of the row p_j^T which forces the zeros in the right location in a_i. For each category c we have a learned dictionary D_c that contains atoms and each atom represents the key features of the samples in each category $\{Y_c \in R^{n \times N_c} : c = 1, \ldots, 4\}$ and the total number of feature samples is represented by $(Y = (Y_1, Y_2, \ldots, Y_4))$ in the dictionary:

$$D = (D_1, D_2, \ldots, D_4) \in R^{n \times N}, \qquad N = \sum_c N_c \tag{12}$$

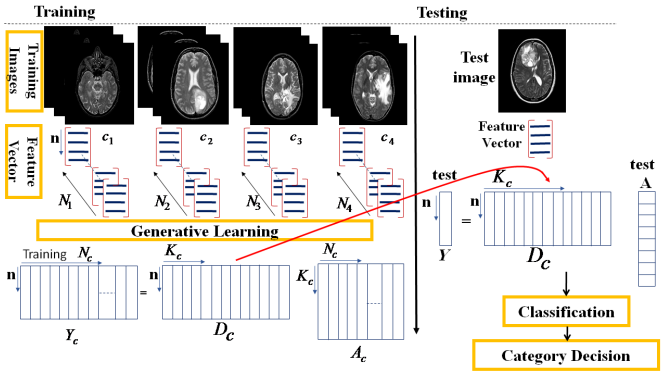

Fig. 2. Schematic illustration of the proposed algorithm. In the training part, (c_1) normal, (c_2) glioma, (c_3) glioplastoma, (c_4) carcinoma and the feature vector represents the topological and texture features of each case. Generative learning illustrates the dictionary learning step. The testing part illustrates the computing of the sparse representation of the test images.

2.3. Classification

The classification step of the proposed algorithm is based on sparse representation. In Sec. 2.2, the dictionary learning step

of each category was explained. To classify the testing data, the algorithm tries to find a match between the testing data Y and the dictionary of specific category D_c. This can be achieved by computing the similarity of the testing data with contents (key features) of the dictionary D_c. Therefore, the sparse representation of the testing data is computed using the individual dictionaries of the all categories (as illustrated in the testing part of Fig. (2)) and then Y is classified as a c^{th} category when appear that Y is more sparse with $D_{c^{th}}$:

$$\|Y - D_c A_c\|_F^2 \leq \epsilon \ , \ \ \|A_c\|_0 = min\{\|A_b\|_0 : b = 1, \ldots, 4\} \tag{13}$$

3. EXPERIMENTAL RESULTS AND DISCUSSION

To explore the advantages of the proposed algorithm compared to the other methods, several experiments have been conducted on diverse medical images. In this paper three medical datasets are used, namely, brain web for simulated brain database [17], brain tumor segmentation database [18, 19], and whole brain atlas [20], and other medical images with brain tumor from the internet. From all databases, 4 classes of images have been collected; 50 normal brain cases (class 1), 50 cases with brain glioma (class 2), 50 cases with brain glioplastoma (class 3), and 50 cases with brain metastatic carcinoma (class 4). Each case has a set of 10 images which make the total number of images for the training set of 4 classes 2000 images (50 cases $\times$ 4 classes $\times$ 10 images of each class = 2000 images). For testing, a 10-fold cross validation is used to evaluate the performance of the classification. Figure (2) shows examples of images from these databases.

Each patch in the dictionary is represented using a feature vector, including topological and textural information. The images are classified as normal case or abnormal case according to their topological properties as explained in Table 1. Then the images are classified furthermore according to the texture features of the abnormality if they exist. In the classification step, two types of classifiers are used (sparse coding classifier and Linear-SVM classifier). Sparse coding classifier performs higher classification accuracy than Linear-SVM classifier (93.7 % versus 88.75 %). Furthermore, the proposed algorithm (using sparse coding classifier) is compared with other classification methods used in [6, 10] after adapting these methods for multi-class classification. In the proposed algorithm, the classification step is obtained by finding the match between the sparse representation of the testing data with the specific dictionary. The performance for multi-class classification (Recall, Precision, Average Accuracy (AA)) are computed by computing the True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) using the algorithm [21]:

$$Precision = \frac{\sum_{c=1}^{C} \frac{TP_c}{TP_c+FN_c}}{C}, \ Recall = \frac{\sum_{c=1}^{C} \frac{FP_c}{TP_c+FN_c}}{C}$$

	c1	c2	c3	c4
c1	0.99	0	0.01	0
c2	0	0.9565	0.0435	0
c3	0	0.0565	0.9435	0
c4	0	0.08	0.06	0.86

Fig. 3. Confusion matrix for all datasets. The average accuracy is 93.75%. Most confusions occur in brain carcinoma. (c_1) Normal, (c_2) Glioma, (c_3) Glioplastoma, (c_4) carcinoma.

Table 2. Classification Evaluation.

Classifier Type	Recall	Precision	AA
Sparse Coding (Proposed)	**92.5**%	**94.87**%	**93.75**%
Han et al. [6]	**92.5**%	90.24%	91.25%
Weiss et al. [10]	90.0%	92.31%	90.0%

$$AA = \frac{\sum_{c=1}^{C} \frac{TP_c+TN_c}{TP_c+TN_c+FP_c+FN_c}}{C}$$

Table 2 illustrates the classification performance of the proposed algorithm using sparse coding classifier better than other classification methods proposed in the literature [6, 10]. From the confusion matrix in Fig. (3), it can be observed that the class 1 (normal) is classified correctly with minimum because the topological feature gives accurate information of the normal and abnormal cases. The errors occurred mainly with the class 4 (carcinoma) with error 0.14 which is classified as class 2 (glioma) and class 3 (glioplastoma) due the textural similarity in T2 MRI images of these cases. Class 3 (glioplastoma) is classified as class 2 (glioma) with error 0.0565 and class 2 (glioma) is classified as class 3 (glioplastoma) with error 0.0435. Totally, 6.25 % of the four classes are classified incorrectly.

4. CONCLUSION

In this paper dictionary learning and sparse coding are proposed for multi-class brain tumor classification. The dictionary is constructed and learned from the topological and texture features of the trained data. Then the learned dictionary is used to classify the testing data. Two types of classifiers are used for classification namely sparse coding and linear-SVM. The sparse coding classifier computes the matching between the sparse representation of the testing data and the corresponding dictionary. The results showed that the sparse representation based classification achieves higher classification accuracy than Linear-SVM based classification technique (93.75 % versus 88.75 %). The proposed algorithm has also been compared to other classification methods, demonstrating the advantages of the method proposed in this paper.

5. REFERENCES

[1] Gladis Pushpa Rathi, V. P., Palani, S. Linear Discriminant Analysis For Brain Tumor Classification Using Feature Selection. International Journal of Communications and Engineering 2012;5(5):130-134.

[2] Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., Ahuja, C. K. Segmentation, Feature Extraction, Multiclass Brain Tumor Classification. Journal of Digital Imaging 2013;26:1141-1150.

[3] Padma, A., Sukanesh, R. Automatic Classification and Segmentation of Brain Tumor in CT Images using Optimal Dominant Gray level Run length Texture Features. International Journal of Advanced Computer Science and Applications 2011;2(10):53-59.

[4] Thiagarajan, J. J., Ramamurthy, K. N., Rajan, D., Spanias, A. Kernel Sparse Models for Automatic Tumor Segmentation. International Journal on Artificial Intelligence Tools 2013; 12:1-12.

[5] Bauer, S., Nolte, L. P., Reyes, M. Fully Automatic Segmentation of Brain Tumor Images using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization. MICCAI 2011;14(3):354-361.

[6] Han, J., Chang, H., Loss, L., Zhang, K., Baehner, F. L.,Gray, J. W., Spellman, P.,Parvin, B. Comparison of Sparse Coding and Kernel Methods For Histopathological Classification of Glioblastoma Multiforme. ISBI 2011;9:711-714.

[7] Selvaraj, H., Thamarai, S., Selvathi, D., Gewali, L. Brain MRI Slices Classification Using Least Squares Support Vector Machine. International Journal of Intelligent Computing in Medical Sciences and Image Processing 2007;1(1):21-33.

[8] Moon, N., Bullitt, E., Leemput, K. V., Gerig, G. Automatic Brain and Tumor Segmentation. MICCAI 2002;LNCS (2488):372-379.

[9] Cocosco,C. A., Zijdenbos, A. P., Evans, A. C. A Fully Automatic and Robust Brain MRI Tissue Classification Method. Medical Image Analysis 2003;7:513-527.

[10] Weiss, N., Rueckert, D.,Rao, A. Multiple Sclerosis Lesion Segmentation Using Dictionary Learning and Sparse Coding. MICCAI 2013:735-742.

[11] Wu, P., Xie, K., Zheng, Y., Wu, C. Brain Tumors Classification Based on 3D Shape. Advances in Intelligent and Soft Computing 2012;160:277-283.

[12] Otsu, N. A Threshold Selection Method from Gray-Level Histograms. IEEE Transaction on System, Man and Cybernetics 1979;9(1):62-66.

[13] Al-Shaikhli, S. D. S., Yang, M. Y., Rosenhahn, B. Medical Image Segmentation Using Multi-level Set Partitioning with Topological Graph Prior. Pacific-Rim Symposium on Image and Video Technology PSIVT Workshops 2013;LNCS(8334):157-168.

[14] Haralick, R. M., Shanmugam, K., Dinstein, I. Texture Feature for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics 1973;SMC-3(6):610-621.

[15] Aharon, M., Elad, M., Bruckstein, A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. IEEE Trancsaction on Signal Processing 2006;54(11):4311-4322.

[16] Duarte-Carvajalino, J. M., Sapiro, G. Learning to Sense Sparse Signals: Simultaneous Sensing Matrix and Sparsifying Dictionary Optimization. IEEE Trancsaction on Image Processing 2009;18(7):1395-1408.

[17] Cocosco, C. A., Kollokian, V., Kwan, R. K-S., Evans, A. C. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. NeuroImage 1997;5(4):S425.

[18] Kaus, M., Warfield, S. K., Nabavi, A., Black, P. M., Jolesz, F. A., and Kikinis, R. Automated Segmentation of MRI of Brain Tumors. Radiology 2001;218(2):586-91. http://www.spl.harvard.edu/publications/item/view/169

[19] Warfield, S. K., Kaus, M., Jolesz, F. A., and Kikinis, R. Adaptive, Template Moderated, Spatially Varying Statistical Classification. Medical Image Analysis 2000;4(1):43-55. http://www.spl.harvard.edu/publications/item/view/209

[20] Whole brain atlas database. http://www.med.harvard.edu/aanlib/home.html

[21] Sokolova, M., Lapalme, G. A Systematic Analysis of Performance Measures For Classification Tasks. Information Processing and Management 2009;45:427-437.