Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 127 (2018) 218–225

Procedia Computer Science

www.elsevier.com/locate/procedia

The First International Conference on Intelligent Computing in Data Sciences

Performance evaluation of feature extraction techniques in MR-Brain image classification system

Mohammed Khalil[a], Habib Ayad[a], Abdellah Adib[a]

[a]*Laboratory of computer sciences FST Mohammedia, Hassan II University of Casablanca, B.P. 146 Mohammedia 20650, Morocco*

Abstract

In this paper, we present a MR-Brain image classification system to classify a given MR-brain image as normal or abnormal. This system first employs three feature extraction techniques namely, Gray-Level Co-Occurrence Matrix (GLCM), Local Binary Pattern (LBP) and Histogram of Oriented Gradient (HOG). The obtained feature vector of each technique is passed through a k-Nearest Neighbor (k-NN) classifier. The resulting dissimilarity measure values of the classifiers are combined then by a fusion operator in order to increase the classification accuracy. Two benchmark MR image datasets, Dataset-66 and Dataset-160, have been used to validate the system performance. A cross-validation scheme is adopted to improve the generalization capability of the system. The obtained simulation results are compared with those ones of the existing methods to evaluate the performance of the presented MR-Brain classification system.

Keywords: MR-Brain image classification, LBP, GLCM, HOG, k-NN, Fusion.

1. Introduction

Magnetic resonance (MR) imaging is a well known medical imaging method that is non-invasive and produces high quality images of the anatomical structures of brain[5]. It provides a very important information for improving the clinical diagnosis and enhancing the healthy brain of the patient. However, high volume of imaging data makes manual methods of detecting diseases more difficult and time-consuming[6]. Hence, developing algorithms that make diagnostic of MR-Brain imaging automated and reliable is very crucial.

Various MR-Brain image classification systems have been suggested in the literature. All of these schemes follow different steps including feature extraction, dimension reduction and classification. Classification techniques include Propagation Neural Network BPNN[8], Forward Back Propagation Artificial Neural Network (FP-ANN) and k-Nearest Neighbor (k-NN)[7], Generalized Eigenvalue Proximal SVM (GEPSVM) classifier[10] and AdaBoost algorithm with

* Corresponding author. Tel.: +212-672-079-141.
E-mail address: medkhalil87@gmail.com

10.1016/j.procs.2018.01.117

random forests[11]. At the feature extraction step, wavelet transforms in their different forms are widely used by researchers because they allow analysis of images at various levels of resolution[7,8,9,10,11]. These forms include Discrete Wavelet Transform (DWT), Discrete Wavelet Packet Transform (DWPT) and Stationary Wavelet Transform (SWT). Generally, the extracted feature vectors can be reduced by applying feature reduction methods such as Principal Component Analysis (PCA) and Probabilistic PCA[7,8,9,11].

Almost all the existing works focus on the classification step and utilize wavelet transforms as the main tool of feature extraction technique. In this paper, we aim to improve the performance of existing MR-classification systems by focusing on feature extraction. We apply three feature extraction methods namely, GLCM[1], LBP[2] and HOG[3] and we utilize k-NN as a classifier. These methods are widely used in several pattern recognition schemes, but they have attracted less attention in the field of MR-Brain image classification. The objective of our idea is to enrich the state-of-the-art methods by evaluating the performance of these three feature extraction techniques. Furthermore, we aim to show how we can improve the classification accuracy by exploiting complementary information offered by each technique. Finally, the presented system will be compared with some existing MR-Brain approaches.

This paper is organized as follows. In the next section, we present the proposed MR-Brain image classification scheme by describing the employed feature extraction techniques and showing how to exploit the fusion operation at the classification level. Section 3 presents the obtained results and some comparisons with some state-of-the-art classification approaches. Finally, section 4 concludes the paper.

2. The proposed MR-Brain image classification system

The proposed MR-Brain image classification scheme consists of three stages:

- Feature extraction: for each image, we apply three feature extraction techniques namely GLCM, LBP and HOG providing three feature vectors.
- Classification: the resulting feature vectors are submitted to three k-NN classifiers providing three dissimilarity measure values.
- Fusion operation: the three dissimilarity measure values are combined by a fusion operator to make a decision on the query image (normal or abnormal).

2.1. Feature extraction

2.1.1. Gray-Level Co-Occurrence Matrix

Gray-Level Co-Occurrence Matrix (GLCM) is a statistical method that considers the spatial relationship among pixels. GLCM is a 2D histogram that calculates how often pairs of pixel separated by a certain distance, d_{occ} occur in an image. Let $I(i, j)$ be an image with size $N \times M$, with L gray levels, $I(i_1, j_1)$ and $I(i_2, j_2)$ be two pixels with gray level intensities x_1 and x_2, respectively. When taking $\Delta i = i_2 - i_1$ in the i direction and $\Delta j = j_2 - j_1$ in the j direction, the connecting straight line has a direction θ which is equal to $\arctan\left(\frac{\Delta j}{\Delta i}\right)$. The normalized co-occurrence matrix $C_{\theta,d}$ is defined as[1]:

$$C_{\theta,d}(x_1, x_2) = \left(Num\left\{\left((i_1, j_1), (i_2, j_2)\right) \varepsilon\, (N \times M) \times (N \times M) \diagup A\right\}\right) \diagup K \tag{1}$$

Here A is a given condition, such as $\Delta i = d_{occ}\sin(\theta)$, $\Delta j = d_{occ}\cos(\theta)$. Num represents the number of elements in the co-occurrence matrix and K is the total number of pairs of pixels. Normally, $d_{occ} = 1, 2$ and $\theta = 0^o, 45^o, 90^o, 135^o$ are used for calculation. Five different texture features are defined using co-occurrence matrix as follows:

$$Energy = \sum_i \sum_j C^2_{\theta,d}(i, j) \tag{2}$$

$$Variance = \frac{1}{N \times M} \sum_i \sum_j \left(C_{\theta,d}(i, j) - \overline{m}\right)^2 \tag{3}$$

$$Entropy = -\sum_i \sum_j C_{\theta,d}(i, j) \log_2\left(C_{\theta,d}(i, j)\right) \tag{4}$$