International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.6, No.5 (2013), pp.39-48
http://dx.doi.org/10.14257/ijsip.2013.6.5.04

Researching on Feature Extraction of Brain CT Image

Ribo Ge[1] and Hailong Shen[2]

[1,2]City Institute
Dalian University of Technology, Dalian 116600, China
ececity@.dlut.edu.cn, shl@.dlut.edu.cn

Abstract

According to grayscale changing characteristics of the brain CT, this paper improves region growth algorithm on the basis of in-depth analysis of brain CT image. In order to get the best combination of the features, the algorithm that extracts the shape feature considering the adjacent parts is proposed based on the idea of chart transferring tree in the tree structure comprehensively. Compared with the traditional extracting algorithms which aim at the whole picture or a single part, the experiments show that the proposed algorithm can obtain a higher diagnostic accuracy.

Keywords: *Shape Features, Image Segmentation, Feature Extraction, Medical Image*

1. Introduction

With the development of medical imaging technology, medical imaging diagnosis and interventional treatment raise to a new level, promoting the development of clinical medicine. How to make full use of the previously confirmed cases and the experience of doctors and the information of patients to enable the computer to help doctors quickly, effectively and correctly diagnose the disease, is the target of computer-aided medical diagnosis system. The accuracy of medical diagnosis depends largely on the interpretation of medical images. Making full use of the medical image information to improve the accuracy of medical diagnosis is the key to computer aided diagnosis.

Medical image segmentation and feature extraction is a key technology in medical image processing, which is a prerequisite for medical image recognition and provides important information for computer aided diagnosis. Shape feature is the core of the image, meanwhile is one of the key information of image recognition; therefore, it is one of the most important tools in computer aided diagnosis research.

Traditional shape feature extracts only a whole picture or single parts, without considering the internal structure of parts and the inter relationship, thus the pertinence is not strong, it is hard to get ideal classification results. In order to overcome the defects of traditional extraction method, this paper proposes the method of feature extraction of adjacent parts, realizes the shape feature extraction of brain CT image based on tree structure, gets the ideal accuracy rate of classification.

2. The Segmentation of Brain CT Image

In view of the complexity and fuzziness of medical image and its oriented on-the-spot application, because of a high price to pay for misdiagnosis, the segmentation accuracy is particularly important. Through the analysis of brain CT images and according to the gray value range of brain tissue, this paper has improved region growing algorithm, divides the

ISSN: 2005-4254 IJSIP

area with organ meanings (such as the lateral ventricles), on this basis, extracts the moment of inertia, circularity shape feature variables based on image shape feature in the segmented region. Such local features can reflect the image information better, and it has more practical significance.

The specific algorithm is as follows:

1) Scan image, and the prior knowledge selects the image the central scotoma（x0,y0） as a seed point.

2) To（x0,y0）as the center, calculate its eight neighborhood gray mean $\overline{\mu}$ and variance σ^2.

3) Determine whether the grayscale neighborhood pixel value is included in the [50,150]. If included in this interval, calculate the similarity R of neighborhood pixel （xi , yi）. According to the growth rule of P, and（xi , yi）and（x0,y0）merge onto the stack.

4) Pixels take newly merged as the seed point（x0,y0）, repeat step 2 until the area is not in the expansion.

5) Repeat steps 1-4, end the whole growth process until the each pixel of image has ownership.

In accordance with the above steps，the interested site (target site) of brain CT image can be achieve automatic segmentation. After regional growth, the peripheral regions are labeled in black in order to facilitate the subsequent feature extraction.

If the seed regions is S: mean $\overline{\mu}$ and standard variance σ^2 are used as characteristic quantity to measure gray consistency. Among them: n is on behalf of the number of pixels in already growth areas, $\overline{x}$ as the gray value, I as the detected pixel gray, The mean $\overline{\mu}$ and the variance σ^2 of S are defined as:

$$\overline{\mu} = \frac{1}{n} \sum_{(i,j)\in R} f(i,j) \tag{1}$$

$$\sigma^2 = \sum_{(i,j)\in R} \left(f(i,j) - \overline{x} \right)^2 \tag{2}$$

Similarity R:

$$R = \left[\frac{\frac{n(n-1)}{n+1} (I-\mu)^2}{\sigma^2} \right]^{\frac{1}{2}} \tag{3}$$

The merger guidelines：

$$\max_S \left| f(x,y) - \overline{\mu} \right| < R \tag{4}$$

3. Shape Features

The following briefly introduces several shape features used in this paper:

Perimeter: area boundary length. A simple shape of the object with the perimeter of a relatively short surrounded pixels of the area, so the perimeter is the outer boundary length around all these pixels .

Area: the number of pixels within the boundary. The calculation formula is as follows:

$$S = \sum_{x=1}^{N}\sum_{y=1}^{M} f\left(x, y\right)$$

Function f(x, y) is the two value image. If the value of 1 represents an object, and 0 represents background, then the area is the number of f(x,y)=1.

The long axis and the short axis: the diameter of the boundary is the distance between two points on the boundary which has most distant way, namely, the length of the line between these two points. This line is called boundary long axis, and the long axis vertical and the longest line which has most distant way to two intersection points' boundary distance is the short axis. Meanwhile the length and the orientation of it is very useful to describe the boundary.

Rectangular degree: it embodies the object on the outer rectangle filling degree; reflects a parameter of an object rectangle degree is fitted factor:

$$R = \frac{S_0}{S_{MER}}$$

Among them, S0 is the area of the object, and SMER is the area of MER. R reflects an object to its filling extent. For rectangular object R has a maximum value of 0.1,for the circular object R value of π /4, and the slender, curved object R' value is smaller. Rectangle fitted factor values in the 0 - 10.

Roundness: expresses the index of the circular degree of object.

$$C = \frac{P^2}{4\pi A}$$

In the equation, P expresses the perimeter, A expresses the area.

In the same area condition, for objects of various shapes, the perimeter of round objects is the shortest. So it can be called the densest shape, circular intensity C=1.With the increase of perimeter concave-convex change degree, perimeter also increases accordingly.

Invariant moments: when the image is the two value image, it is available to describe the shape feature of some areas in an image:

The probability density distribution function is regular matrix order (p+q) of two-dimension continuous random function of f(x,y) (also known as geometric moments) . m_{pq} is defined as:

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f\left(x, y\right) dxdy \; m_{pq}$$

The regular matrix of two-dimensional (M * N) digital image f (m, n) is defined as:

$$m_{pq} = \sum_{m=1}^{M}\sum_{n=1}^{N} m^p n^q f\left(m, n\right) \; m_{pq}$$

Complete set of regular matrix order n include all m_{pq} meet the $p+q \le n$, thus there are $\frac{1}{2}(n+1)(n+2)$ elements. Monomial product $x^p y^q$ is the base function of the matrix definition.

Central matrix:

$$\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(x-\overline{x}\right)^p \left(y-\overline{y}\right)^q f\left(x,y\right) dxdy$$

In the equation, $\overline{x} = m_{10} / m_{00}$, $\overline{y} = m_{01} / m_{00}$.

For the digital image, available summation instead of integral:

$$\mu_{\mathrm{pq}} = \sum_{m=1}^{M} \sum_{n=1}^{N} \left(m-\overline{m}\right)^p \left(n-\overline{n}\right)^q f\left(m,n\right)$$

(p+q) order normalized central matrix:

$$\eta_{pq} = \mu_{pq} / \mu_{00}^{r}$$

In the equation, $r = 1 + {}^{(p+q)}\!/_{2}$, $p,q = 1,2,3\cdots$

The following are the seven invariant matrices:

$$\phi_1 = \eta_{02} + \eta_{20}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]$$
$$+(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

$$\phi_6 = (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$
$$+4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]$$
$$+(\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

When the area is translated, rotated and scaled, these invariant matrices are invariant. Invariant matrix have good shape feature.

4. Shape Feature Extraction

Brain disease is relatively more, in order to accurately analyze the brain CT images, this paper makes positive anomaly judgment for a disease (disease cerebral hemorrhage). Shape characteristics of the target site is extracted from the image of a layer of brain CT image, and this paper uses SEE5, the extension matrix and RBF neural network classifier to classify and test the extracted data.

4.1. Basic Idea

After image segmentation and labeling, according to the a priori knowledge, shape feature extraction for the area of interest of image (*i.e.*, target site) can be extracted. For example: Figure 1 is a brain CT image of normal, Figure 2 is main symptoms of cerebral hemorrhage. In normal circumstances, as shown in Figure 2 (a), deformation of lateral ventricle lateral ventricle is caused by the extrusion of cerebral hemorrhage, so extract the feature of the lateral ventricle.

Another phenomenon will appear: as shown in Figure 2 (b), cerebral hemorrhage symptoms did not cause deformation of the lateral ventricle. In view of the situation, there is a need to consider its adjacent site. The increased number of the target site（the lateral ventricle）in adjacent areas caused by cerebral hemorrhage symptoms, extraction of the number of adjacent area is to provide rich data for accurate diagnosis.

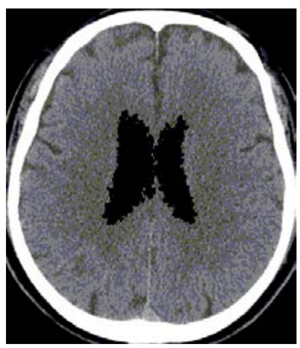
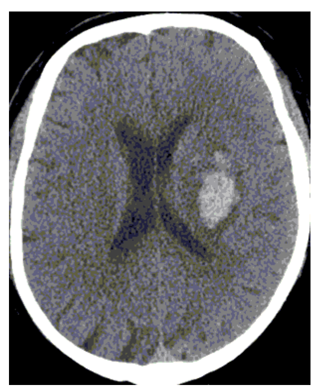


Figure 1. Normal Brain CT Image

(a) Extrusion Deformation

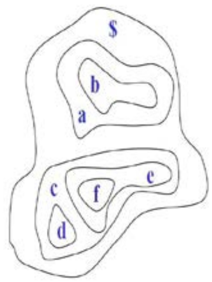


(b) Without Deformation

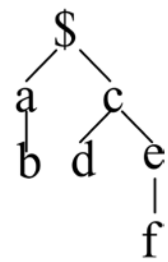
Figure 2. Abnormal Brain CT Image

Therefore, according to the idea of figure turn to tree, this paper realizes a comprehensive method of extracting shape feature of adjacent site.

The basic idea: according to the relationship the closed area contains, the segmented images gradually create tree. For example: Figure 3 (a) after segmentation, the outer contour \$ of graphics is the root node of the tree type structure, the contour of \$ contains a and c, two closed areas as the root node of \$, a also contains closed region b, so b becomes the child nodes of the a.



(a) The Image after Segmentation

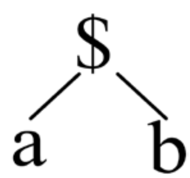


(b) The Tree Structure

Figure 3. Structure the Tree Structure with \$ as the Root Nnode

After the establishment of tree structure, each node in the tree structure represents a closed area, father and son nodes represent inclusion relation, and brother node represents a parallel relationship. The position relationship (including and adjacent) between various organs of brain CT images can be clearly reflected by tree structure to bring great convenience for computer aided diagnosis.

(a)The Labeled Brain CT Image



(b)After Conversion of the Tree structure

Figure 4. Construct the Tree Structure of the Brain CT Image

Figure 4 as an example: brain CT image after segmentation and labeling, as shown in Figure 4 (a), divided into \$area, a area and B area. According to the idea of figure turn to tree, make mark brain CT image into a tree structure Figure 4 (b). So, according to Figure 4 (b) tree structure is shown, in accordance with a priori knowledge of medical science, we can extract interest site (in this case lateral ventricle site i.e. a node in tree structure), traverse the tree structure, and determine the number of adjacent parts, and extract shape feature extraction for the area of interest of image (*i.e.*, target site).

Therefore, according to the translation, rotation and scaling invariance, this paper selects seven moment invariants, compactness, moment of inertia, the number of adjacent regions and the area ratio of father and son nodes as the shape feature of the target site. In this case, it is the lateral ventricle locations *i.e.*, a node in the tree structure of Figure 4 (b).

4.2. The Brain CT Image Feature Extraction Algorithm

Design steps: first, according to the characteristics of the brain CT image, selects improved region growing algorithm. Secondly, from top to bottom, left-to-right orders to mark the divided region. Again, according to the idea of figure turn to tree, picture turns into a tree structure. Finally, traverses the tree to extract the target site characteristics.

The specific algorithm:

- Image segmentation
- Region labeling, acquiring a target site number
- Figure turn to tree
- Traversing the tree
  - Looking for interested parts numbers to get the number of Sibling nodes
  - Extracts relevant features of the target site
- End

5. Experimental Results and its Analysis

The brain CT images provided by hospital are always 12 layers and the eighth layer is chosen. We used 412 images lastly which included 203 normal images and 209 abnormal images for this test.

According to the algorithm we can extract the shape characteristics of the target site, and its feature is as follows: normal data is relatively concentrated, abnormal data is more dispersed, as we can see in the Figure 5. This phenomenon is mainly because the normal brain CT images performance are more similar while the abnormal brain CT images performance vary greatly because of different diseases.

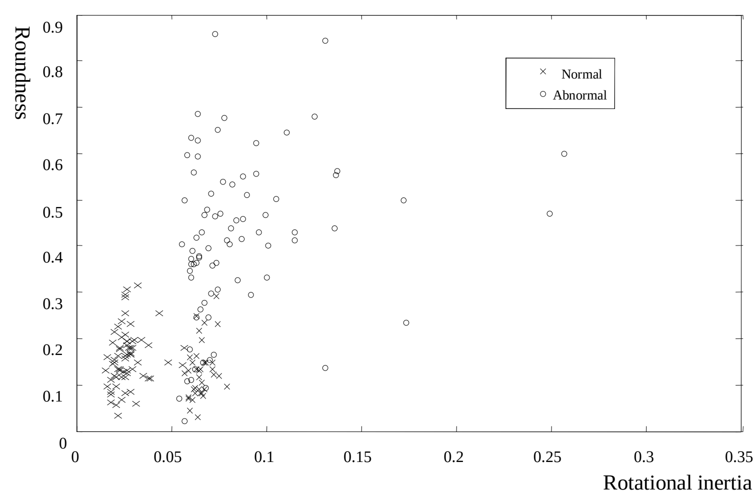


Figure 5. Experimental Results

Select SEE5, extension matrix and RBF neural network methods to test the head CT images data. The experimental procedure is as follows:

1. Data preprocessing: normalized the original feature data to [0, 1].

2. Cross-validation: these original images were randomly divided into 10 samples with the same number and non-overlapping subset. Take a subset in turn as a test set, and then the remaining nine subsets are merged as training set.

3. Training: used SEE5, Extension matrix and the RBF neural network to train the training set.

4. Test: used the training model and parameters to test the test set, then we can get the test accuracy.

Table 4.1. Classification Results of the Brain CT

Classification model	Training accuracy	Test accuracy
SEE5	93.7%	89.5%
Extension Matrix	92.6%	85.6%
RBF NN	90.2%	86.3%

Table 4.2. The Main Features Used in Classification of Brain CT

Classification model	Main features
SEE5	Roundness, rotational inertia, and the number of the adjacent region, matrix one, area ratio
Extension Matrix	Roundness, rotational inertia, and the number of the adjacent region, matrix one , matrix two
RBF NN	All

Table 4.3. Classification Results of the Whole Picture

Classification model	Training accuracy	Test accuracy
SEE5	63.7%	59.5%
Extension Matrix	62.6%	55.6%
RBFNN	60.2%	56.3%

Table 4.4. Feature Extraction Classification Results

Extraction style	Training accuracy	Test accuracy
Whole picture	63.7%	59.5%
Single site	82.6%	74.9%
Tree structure	93.7%	86.8%

Analysis of the experimental results:

From the Table 4.4 we can find, the classification result from the whole picture is not satisfactory while its average classification accuracy is 59.5%. Use the single-site feature, the classification results have been significantly improved to 74.9%. However, the cost of the medical misdiagnosis is very large; it still cannot meet the application requirements. From the experimental results it can be seen, with comprehensive consideration of the multi-part and extract features with tree structure, we can obtain significantly improvement comparing with the original algorithm. The average classification accuracy reached 86.8%.

As can be seen in Table 4.1, with a variety of classification methods, the average classification accuracy can achieve a relatively high level. It indicates that assembling the algorithm with classical shape features can describe the target shape fully. This method improved the classification of the image accuracy, and has a good practicability.

It can be seen from the experimental results, the features extracted from the above algorithms can describe the shape of the object fully, and the high classification accuracy is helpful to improve the accuracy of medical image recognition, which is basic for medical diagnosis.

Because the structure of the 8 floor in brain CT is relatively simple, image segmentation preprocessing workload is relatively small, so we can have high classification accuracy. But for the complex level structure, such as the first layer of the brain CT, image segmentation complexity greatly increases, the accuracy of the image segmentation is also relatively decreased, affecting the precision of classification. Therefore, the effect of pretreatment of the picture is not ideal, the processing technology is still relatively lacking, and the program should be further improved.

By Analyzing the first layer of the brain CT image we can find, the classic picture features needs to be understood from multiple perspectives and to mine its role. For example, we can use the characteristics of the brain CT image symmetry, combined with the symmetry characteristics of the image to make block, calculate the corresponding two block's similarity (Euclidean distance) and use it as the image's feature, correct classification rate can reach 85% which is 9 percentage points improved. Therefore, combination of the image symmetrical, gray and texture features and making the best combinations will make higher classification accuracy.

References

[1] Z. F. Zhao and Y. S. Zhang, "Summary of medical image segmentation", East China Shipbuilding Institute (Natural Science), vol. 17, no. 3, **(2003)**, pp. 43-47.

[2] W. C. Luo, "Research on the method of image segmentation and its comprehensive integration", Sun Yat-sen University: Ghuang Zhou, **(2001)**.

[3] J. S. Duncan and A. Yache, "Medical image analysis: progress over two decades and the challenges ahead", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, **(2000)**, pp. 85-106.

[4] S. A. Hojjatoleslami and J. Kittler, "Region growing: A new approach", IEEE Trans. Image Process, vol. 7, **(1998)**, pp. 1079-1084.

[5] N. Patel and M. Zaveri, "3D Facial Model Construction and Expressions Synthesis using a Single Frontal Face Image", IJCG, vol. 1, no. 1, **(2010)** November, pp. 1-18.

[6] H. Wang, "Fast Image Fractal Compression with Graph-Based Image Segmentation Algorithm", IJCG, vol. 1, no. 1, **(2010)** November, pp. 19-28.

[7] I. Kanta Maitra, S. Nag and S. Kumar Bandyopadhyay, "Identification of Abnormal Masses in Digital Mammography Images", IJCG, vol. 2, no. 1, **(2011)** May, pp. 17-30.

[8] H. Al-Nuaim (DSc), "Nouf Abukhodair. A User Perceived Quality Assessment of Lossy Compressed Images", IJCG, vol. 2, no. 2, **(2011)** November, pp. 23-36.

[9] C.-C. Chang, Y.-Z. Wang and C.-S. Chan, "An Efficient Probability-based t out of n Secret Image Sharing Scheme", IJSIP, vol. 2, no. 1, **(2009)** March, pp. 71-78.

Authors

Ribo Ge received his Master's degree in Computer Science and Technology from Dalian University of Technology in Dalian, China. He is currently an adjunct professor in City Institute, Dalian University of Technology in Dalain, China. His research interest is mainly in the area of Computer Application, Algorithm Design and Theory.

Hailong Shen received his Master's degree in Management Science and Engineering from Dalian University of Technology in Dalian, China. He is currently a lecture in City Institute, Dalian University of Technology in Dalain, China. His research interest is mainly in the area of Computer Science and Information Technology.