

# RECOGNITION OF IMAGE WITH NATURAL TEXTURES BASED ON LEARNING OF INFORMATION AUGMENTATION

Cheng Xian-Yi<sup>1,2</sup>, Yuan Xiao-Hua<sup>1</sup>, Li Shu-Qin<sup>1</sup> and Xian De-Shen<sup>1</sup>

<sup>1</sup>*Department of Computer Science, NanJing University of Science & Technology, Nianjing 219004, P.R.china*

<sup>2</sup>*State Key Laboratory for Novel Software Technology at Nanjing University, Nianjing, 210093, P.R.China*

<sup>3</sup>*Computer Science & Communication Engineering Institute of Jiangsu University, Zhenjiang, 212013, P.R. China*

**Abstract:** The efficiency of pattern recognition depends heavily on that if feature extraction and selecting are effective. Complicated image such as medical image and remote sensing image, belong to image with natural textures, this kind of image is always of high resolution, with many layers of gray degree, and a very intricate shape structure. Because there are no obvious shapes, but only distributions of some gray degrees and colors in these images, so for them, there are no good methods yet for feature extraction and region recognition. In this paper, based on information augmentation and kinetics, we present a learning algorithm, which can be used to do region classification of the above-mentioned images with natural textures. We applied our algorithm to recognition of image with natural textures and obtained a good result.

**Key words:** pattern recognition; information entropy augmentation; machine learning; NN

## 1. INTRODUCTION

In many medical images, such as cardia and cerebrum, they are neither any obvious edge nor clear boundary between different tissues; we call these images natural textures. For this kind of medical images, it is very difficult to do region recognition. For example, when classifying human cerebrum

tissues in clinical application by computer, features used mostly are ectocinerea, alba, cerebrorachidian liquid, pallium and background etc. In image of Magnetic resonance imaging( MRI), although different tissues and background may have different gray degrees, each tissue doesn't belong to a single gray degree but belong to a gray degree scope, and the gray degree distributions of cerebrum tissues can be nearly taken as gaussian distribution, and further more, the distributive functions of the gray degree of these tissues overlay each other, we can see this from Fig. 1.(data come from document[1], p93).

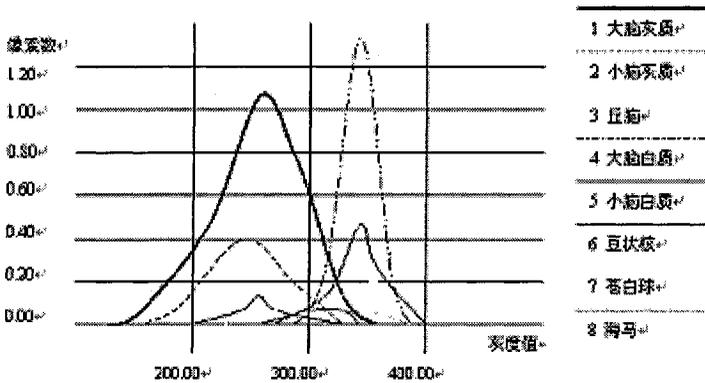


Fig. 1 Gray degree distribution of human cerebrum tissues in MRI

All that mentioned above, we classify these tissues neither by gray degree thresholds nor by lower feature, which based statistics, such as color, texture, shape, etc, because these features are not competent for describing image with natural textures. So in this paper, to aim at the difficulty of region recognition of image with natural textures, based on the information augmentation and kinetics, we put forward a learning algorithm, which transform a minimizing problem of information entropy augmentation into a learning process of dynamic system. Applied our learning algorithm in interesting regions recognition of human cerebrum, we got a satisfied result.

## 2. INFORMATION AND INFORMATION AUGMENTATION

Information theory, also called statistical communication theory, firstly set up by Shannon to solve problems from information transfer (communication) process, it is a theory to study the generation, obtaining, measuring, transformation, transfer, processing, recognizing and application

of information. Information entropy is a measure of the uncertain degree in information<sup>[2,3]</sup>. The related concept, we summarize as follow:

(1)Self information of event. If S represent a group of event marked as  $E_1, E_2, \dots, E_n$ , their corresponding probabilities are  $P(E_j) = p_j, 0 \leq p_j \leq 1$ , and the regularization constraints on  $p_j$  is

$$p_1 + p_2 + \dots + p_n = 1 \quad (1)$$

Then the self information of event  $E_j$  is defined as  $I(E_j) = -K \ln p_j \quad i = 1, 2, \dots, n$ , where  $K = \log_2 e$ .

(2)Entropy of S. Entropy of S is the statistical mean of S's Self information, we note it as  $H(s)$

$$H(s) = E(I) = -K \sum_j p_j \ln p_j \quad (2)$$

some remarks related to (2) are:

- if  $H(s) = 0$ , then there exists only one possibility and no uncertainty.
- if the occurring possibility of n events are the same, namely, to event  $E_j, P(E_j) = 1/n$ , then  $H(s)$  achieve it's maximum  $-K \ln n$ , and the corresponding system is with the largest uncertainty.

· If  $P(E_j)$  closing to each other, then  $H(s)$  is large, and vice versa.

(3)Information augment. Rewriting formula (2) as formula (3)

$$H(s) = \sum_j p_j (-K \ln p_j) = \sum_j p_j I(E_j) \quad (3)$$

Then we have

$$I(E_j) - I(E'_j) = -K \ln p_i + K \ln p'_j = K \ln \left( \frac{p_j}{p'_j} \right) \quad (4)$$

Averaging formula (4), and take the result formula (5) as the measure of mean changes of information (we call it information augmentation)

$$VI(S, S') = K \sum_j p_j \left( \frac{p_j}{p'_j} \right) \quad (5)$$

Assume that  $p_j, p'_j$  meet(1), then it can be proved that  $VI(s) \geq 0$ , and that only when S is of an uniform distribution, equation (5) come into exist.

### 3. LEARNING BASED INFORMATION AUGMENTS

Let  $f(q)$  be an actual pattern distributive function, and  $\tilde{f}(q)$  be the guessed distributive function generated by the system, if we take information augmentation as the measure of the similarity degree between the two

distributive function above, then we can denote information augmentation as below (might as well let  $K=1$ ):

$$VI(S, \tilde{S}) = \int f(q) \ln \left( \frac{f(q)}{\tilde{f}(q)} \right) d^n q \geq 0 \quad (6)$$

Where  $f(q)$  and  $\tilde{f}(q)$  satisfy two regularization constraints in (7)

$$\int f(q) d^n q = 1 \quad \text{and} \quad \int \tilde{f}(q) d^n q = 1 \quad (7)$$

We can rewrite (6) as (8)

$$VI(S, \tilde{S}) = \int f(q) \ln f(q) d^n q - \int f(q) \ln \tilde{f}(q) d^n q \quad (8)$$

Because  $f(q)$  having an invariant value, so we can get the minimum of  $VI(S, \tilde{S})$  as long as maximizing (9)

$$\int f(q) \ln \tilde{f}(q) d^n q = \max \quad (9)$$

Let  $q = (q_1, q_2, \dots, q_n)$ , then the problem left is how to estimate  $\tilde{f}(q)$  by measuring all the moments in (10)

$$\langle q_i \rangle, \langle q_i, q_j \rangle, \langle q_i, q_j, q_k \rangle, \langle q_i, q_j, q_k, q_l \rangle \quad (10)$$

In accordance with maximum information(entropy) theory, the estimating formula of  $\tilde{f}(q)$  is of form as

$$\tilde{f}(q) = N \exp(\sum_j \lambda_j q_j - \sum_{ij} \lambda_{ij} q_i q_j + \sum_{ijk} \lambda_{ijk} q_i q_j q_k - \sum_{ijkl} \lambda_{ijkl} q_i q_j q_k q_l)$$

and because  $\tilde{f}(q)$  is regularization constrained under space of  $q$ , the odd moments need to be assumed to be zero, such we have

$$\tilde{f}(q) = N \exp(-\sum_{ij} \lambda_{ij} q_i q_j - \sum_{ijkl} \lambda_{ijkl} q_i q_j q_k q_l) \quad (11)$$

formula (11) can be realized by using a NN with 4 layer, then  $\lambda_{ij}$  and

$\lambda_{ijkl}$  are the corresponsive

connective weights, and then solving (9) is transformed to solving for  $\lambda_{ij}$  and  $\lambda_{ijkl}$ . Below we will

deduce the learning process in which  $\lambda_{ij}$  and  $\lambda_{ijkl}$  are solved:let

$$N = \exp(-\lambda), V_{j'} = \begin{cases} q_i q_j & \text{if } j' = ij \\ q_i q_j q_k q_l & \text{if } j' = ijkl \end{cases}, \text{ and}$$

$$\lambda_{j'} = \begin{cases} \lambda_{ij} & \text{if } j' = ij \\ \lambda_{ijkl} & \text{if } j' = ijkl \end{cases}$$

then (11) changes to (12)

$$\tilde{f}(q) = \exp(-\lambda - \sum_j \lambda_j V_j) \quad (12)$$

multiply the left side of (9) by  $-1$ , and according to (11), there is

$$W = - \int f(q) \ln \tilde{f}(q) d^n q = \lambda \int f(q) d^n q + \int f(q) \sum_j \lambda_j V_j d^n q,$$

and because of (7), there is

$$W = \lambda + \int f(q) \sum_j \lambda_j V_j d^n q \tag{13}$$

where  $\lambda_j$  was controlled by the strategy of gradient evolvment, namely  $\lambda_j$  subjects to (14)

$$\dot{\lambda}_j = -\gamma \frac{\partial W}{\partial \lambda_j} \tag{14}$$

From (7) and (12), we get

$$e^{-\lambda} \sum_j \exp(-\sum_k \lambda_k V_k) = 1,$$

and consequently there is

$$\lambda = \ln \sum_j \exp(-\sum_k \lambda_k V_k) \quad \text{or}$$

$$\lambda = \ln \int \exp[-\sum_j \lambda_j V_j] d^n q \tag{15}$$

put (15) into (13) and take the derivation of (13) in direction of  $\lambda_j$ , we can get PDEs as (16)

$$\frac{\partial W}{\partial \lambda_j} = - \left\{ \left( \int \exp[-\sum_j \lambda_j V_j] d^n q \right)^{-1} \left[ V_j \exp[-\sum_j \lambda_j V_j] d^n q \right] + \int V_j d^n q \right\} \tag{16}$$

It is apparently that the expression in the big bracket in (16) can be interpreted as the mean value of  $V_j$  on the distributive function of  $\tilde{f}(q)$ , namely  $\{.....\} = \langle V_j(q) \rangle_f$ , and the second term of the right side is the same mean value namely  $\langle V_j(q) \rangle_f$  of the distributive function prescribed by outside, then (14) can be rewritten with a more simply form as (17)

$$\dot{\lambda}_j = \gamma (\langle V_j(q) \rangle_f - V_j(q) \rangle_f) \tag{17}$$

Because each mean value is just its corresponding measure moment, so as concerned to the distributive function given by experiment, to decide the mean value of  $V_j$  is very simply. In practice, we can do that in this way: assuming that  $f(q)$  is an uniformly distributive function of an incoming signal, and that these signal arrive at discrete time  $\tau$ , then each incoming signal can be described as a corresponding  $q_\tau$ . When there have been L surveyed results, we will have:

$$\langle V_j(q) \rangle_f = \frac{1}{L} \sum_{\tau=1}^L V_j(q_\tau) \rightarrow \frac{1}{T} \int_0^T V_j[q(\tau)] d\tau.$$

Because time consumed to compute  $\langle V_j(q) \rangle_{\tilde{f}}$  is too more if using the above formula, so alternately, we make use of the properties of a steady stochastic process, and we can do so just because there is

$$V_j(q) = \langle V_j(q) \rangle_{\tilde{f}} + p_j(t)$$

Where  $p_j(t)$  is a fluctuate variant that satisfy  $\langle p_j(t) \rangle = 0$ . From all above, (17) become

$$\dot{\lambda}_j = \gamma(V_j(q) - \frac{1}{T} \int_0^T V_j[q(\tau)]d\tau) \tag{18}$$

The formula (18) Shows that to solving for  $\lambda_j(t)$ , we must use  $\lambda_j(t-1)$  decided in the before step, so the process solving for  $\lambda_j(t)$  become a learning process.

#### 4. LEARNING ALGORITHM

We summarize our learning algorithm as follow:

step 1. Initial feature space  $q = (q_1, q_2, \dots, q_n)$ , learning time  $T$ , learning step length  $\Delta t$ , learning parameter  $\gamma$ ,  $t = 0$ ,  $\lambda_i(t) = 0, i = 1, \dots, n$ , actual pattern distributive function  $f(q)$ , error threshold  $\varepsilon$ , the largest learning step length  $T'$ , and let the initial learning step length  $\Delta t = 0.005$ .

step 2. Compute  $V_j(q), \frac{1}{T} \int_0^T V_j[q(\tau)]d\tau, j = 1, \dots, n$ .

step 3. Update  $\lambda$ :

$$\lambda_j(t+1) = \lambda_j(t) + \gamma \{V_j(q) - \frac{1}{T} \int_0^T V_j[q(\tau)]d\tau\}, j = 1, \dots, n.$$

step 4. Compute  $\tilde{f}(q)$  according to (11).

step 5. If  $|f(q) - \tilde{f}(q)| < \varepsilon$ , then Learning succeeds and go to step 9.

step 6. If  $t \leq T$ , then set  $t = t + 1$ , and go to step 2.

step 7. If  $t > T$  and  $\Delta t \leq T'$ , then let  $\Delta t = \Delta t + 0.005, t = 0$ , and go to step 2.

step 8. If  $t > T, \Delta t > T'$  and  $\gamma \leq 1$ , then let  $\Delta t = 0.005, t = 0$ , and go to step 2,

else learning process fail.

step 9. end.

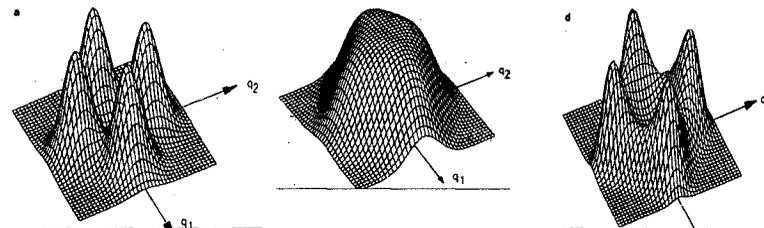


Fig. 2 running samples of learning algorithm

## 5. CONCLUSIONS

In this paper, by utilizing the information entropy augmentation and dynamic process, we transformed a complicate recognition process into a learning process of the NN connecting weight, so solved the recognition problem of image with natural textures.

## ACKNOWLEDGEMENTS

This work was supported by Education Office of Jiangsu province,China under Grant 02KJD520004.

## REFERENCES

- 1 Luo shu-qian and Zhou guo-hong, Medical image processing and analysis. Science press,Beijing, China, 2003 .
- 2 Chang Jiong. Information theoretical foundation[J].Tsinghua University Press, Beijing,China, 2001.
- 3 Haken. Harmonizing. Computer and cognizing - the from above to below method of NN [J]. Tsinghua University Press, Beijing, China, 1994:78-79.