WRITER IDENTIFICATION BASED ON HYPER SAUSAGE NEURON

Samsuryadi, Siti Mariyam Shamsuddin and Norbahiah Ahmad

Universiti Teknologi Malaysia, Malaysia, syamsuryadi@unsri.ac.id

{mariyam, bahiah }@utm.my

ABSTRACT. This paper proposes biomimetic pattern recognition (BPR) based on hyper sausage neuron (HSN) and applies it in writer identification. HSN is used to cover the training set. HSN's coverage can be seen as a topological product of a one-dimensional line segment and an n-dimensional supersphere. The feature extraction is moment invariants such as united moment invariants (UMI) and aspect united moment invariants (AUMI). The experiments result show that AUMI-HSN method is more effective than UMI-HSN method for identifying the authorship of handwriting.

Keywords: biomimetic pattern recognition, hyper sausage neuron, writer identification, united moment invariants, aspect united moment invariants

INTRODUCTION

Research on handwriting analysis based on the identification of the author's point of view in the last ten years experienced a significant development, particularly in forensic applications. A writer identification system aims to search a document legal ownership of a person against a large database with a sample of the author's handwriting recognition (Bulacu & Schomaker, 2007). Special image-making is done based on the features captured from each individual's handwriting. The final decision made by forensic experts to determine the identity of the author sample in question.

One of the problems of identification for purposes of the authors often appear in the court of justice in determining whether a conclusion about the authenticity of the document. This also applies in some institutions that analyze the text of former writers, and identification of various authors who took part in the preparation of the manuscript. The significant results from recent years in the field of handwriting recognition makes it possible to bring this significant answers to specific problems.

At this time, many researchers have used statistical decision model in identify the writer from the handwriting samples. Pattern classification used to determine the pattern without using some previous knowledge of the relationship between the samples in the same class. This differs from the human function.

Human being recognizes things individually by finding the commonalities between things in the same class. This is done by assuming that the sample points of the same class in the feature space would be continuous and recognizable characters. Hence, recognition of a certain class of objects is important, the analysis and cognition of the "shapes" of the infinite point sets constituted by all the objects in feature space. This concept is called biomimetic pattern recognition (BPR) by Wang Shoujue (Shoujue, 2003). BPR concept is incorporated into writer identification for identifying authorship of handwriting (Samsuryadi & Shamsuddin, 2010).

This paper focuses on hyper sausage neuron (HSN) for writer identification. Firstly, some handwritings are extracted through united moment invariant (UMI) (Yinan, et.al, 2003) and aspect united moment invariant (AUMI) techniques (). Secondly, HSN classifier is used to identify the features obtained at the first step. The experiments of writer identification is implemented to demonstrate learning ability and the correct rate of AUMI-HSN and UMI-HSN methods.

WRITE IDENTIFICATION BASED ON HSN

Biomimetic Pattern Recognition (BPR)

In the real world, every one finds one by one similarity between things in the same class. If there are two samples belong to the same class, the differences between them should gradually change. So there must be a sequence of gradual changes between the two samples. Principle of continuity between homologous samples in feature space is called the principle of homology-continuity (PHC) (Shoujue & Xingtao, 2004). PHC can be described in mathematical formulas: suppose that point set *A* includes all samples in the class *A* in feature space. If $x, y \in$ A and ε> 0 are given, there must be set B:

$$\mathrm{B} = \{x_1 = x, ..., x_{n-1}, x_n = y \mid \rho(x_i, x_{i+1}) < \varepsilon,\ \forall i \in [1, n-1], i \in N\} \subset \mathrm{A} \tag{1}$$

It is a kind of prior knowledge of sample distribution in the BPR to improve the cognitive ability, then BPR intends to find the optimal covering of samples in the same class. The basic step of BPR is to analyze the relation between training samples of the same class in the feature space, which is made possible through the PHC of sample distribution (Jiang, at. al., 2009).

Cover Neuron

HSN is as the basic covering unit of the training set. HSN's coverage in high dimensional space, which constructs a sausage like shape in feature space for covering the distribution area of the sampling points in the same class, (Shoujue & Xingtao, 2004). The HSN covering can be seen as a topological product of a one-dimensional line segment and an two-dimensional supersphere (Xu & Wu, 2010).

Cover process

Let $A = \{A_1, A_2, ..., A_n\}$, is the samples points of the training set and one sample denoted $A_i = (a_{i1}, a_{i2}, ..., a_{il})$, where $i = 1,2,...,n$ and l is dimension of the feature space or number of features.

The construction steps of HSN for writer identification are as follows:

Step 1. Calculate the Euclid distance every two points in the A, find two points with the shortest distance, denoted B_{11} and B_{12}. L_1 is segment line $\overline{B_{11}B_{12}}$. HSN covers B_{11} and B_{12} is denoted as H_1, and it coverage is C_1 :

$$C_1 = \{X \mid \rho(X, L_1) \le k\}, X \in R^n\} \tag{2}$$

$$L_1 = \{Y \mid Y = \alpha B_{11} + (1-\alpha) B_{12}, \alpha \in [0,1]\} \tag{3}$$

where $\rho(X, L_1)$ is the distance between the point X and the covering unit L_1.

Step 2. Let $U_1 = S - \{B_{11}, B_{12}\}$. Find point in U_1 is the nearest to B_{12}, denoted as B_{13} and make the second segment line $\overline{B_{12}B_{13}}$, denoted as L_2. HSN covers B_{12} and B_{13} is denoted as H_2, and it coverage is C_2 :

$$C_2 = \{X | \rho(X, L_2) \le k\}, X \in R^n\} \tag{4}$$

$$L_2 = \{Y | Y = \alpha B_{12} + (1-\alpha)B_{13}, \alpha \in [0,1]\} \tag{5}$$

Step i. Delete remaining points which are included in $C_1, C_2, ..., C_{i-1}$. Find point $B_{1(i+1)}$ in the remaining points,which is nearest to B_{1i} denoted line segment $\overline{B_{1i}B_{1(i+1)}}$, is as L_i. HSN covers B_{1i} and $B_{1(i+1)}$ is denoted as H_i, and it coverage is C_i :

$$C_i = \{X | \rho(X, L_i) \le k\}, X \in R^n\} \tag{5}$$

$$L_i = \{Y | Y = \alpha B_{1i} + (1-\alpha)B_{1(i+1)}, \alpha \in [0,1]\} \tag{6}$$

The above algorithm is terminated, if all the points in A have been covered.

Finally we have (n-1) HSNs, and the covering area of training samples in this case is the union set of the areas by these neurons:

$$C = \bigcup_{j=1}^{n-1} C_j \tag{7}$$

In this study, we adopted $k = \beta D_{ij}$, where D_{ij} is the distance between A_i, A_j (Xu & Wu, 2010). β is in the range of [0.30,0.75].

Identifying Algorithm

Calculate the distance ρ_i between sample point A for identifying and the union C_i of class i $(i = 1,2,...,q)$ and ρ_i was defined as formula (8).

$$\rho_i = \min_{1 \le j \le M_i} D_{ij} \tag{8}$$

where D_{ij} was the minimum distance from A to the complex geometrical body C_j $(j = 1,2,..., M_i)$ of union C_i.

Calculated each ρ_i for A. Finally the testing sample A would be classified to the class which corresponding to the least ρ_i namely,

$$r = \arg\min_{1 \le i \le q} \rho_i \tag{9}$$

RESULT AND DISCUSSION

In this paper, the handwriting data are obtained from IAM database. We choose 10 persons with 10 words were selected and each word was made for 10 times (all 1000 samples). We use two feature extraction methods such as united moment invariants (UMI) and aspect united moment invariants (AUMI) to show that BPR is not relied on certain feature extraction method.

For each of 10 persons (writer) has 20 training samples (4 words x 5 repetition), and 25 testing samples (5 words x 5 repetition). Each training samples is used to training the neurons of BPR model for each class, thus each cover set of the 10 persons has 19 HSNs. The experiment result in percentage for beta value 0.30 as far as 0.75 can be showed in Table 1.

Table 1. Percentage Result for Each Writer and Beta Value Based on AUMI-HSN

Writer	Beta value				
	0.30	0.40	0.50	0.60	0.75
W1	56	92	100	100	100
W2	84	92	92	96	100
W3	32	48	64	76	84
W4	52	60	88	88	100
W5	20	40	72	84	96
W6	76	92	92	96	96
W7	64	76	84	88	96
W8	28	52	68	80	92
W9	48	60	84	96	100
W10	20	44	60	76	100
Average	48.00	65.60	80.40	88.00	96.40

Based on Table 1, W1 with beta value 0.30 can be identified 14 samples from 25 samples (56%), 92% (23/25) for beta value 0.40, and so on. The best average result of identifying writer from 10 writers in beta value 0.75 is 96.40%. We can see beta value has influence to identify the authorship of handwriting.

We do the same way for UMI-HSN with 10 writers, 20 training samples and 25 testing samples and the best average result in beta value 0.75 is 88.00%, detail result shows in Table 2.

Table 2. Percentage Result for Each Writer and Beta Value Based on UMI-HSN

Writer	Beta value				
	0.30	0.40	0.50	0.60	0.75
W1	24	64	80	96	96
W2	36	48	84	96	96
W3	8	20	32	76	88
W4	32	72	80	84	88
W5	24	36	64	76	84
W6	0	4	20	28	40
W7	40	48	56	68	88
W8	32	48	84	92	100
W9	32	84	92	100	100
W10	32	44	64	88	100
Average	26.00	46.80	65.60	80.40	88.00

Besides experiment above, we do the other training samples and testing samples to show the performance of the method. For instance, UMI(30,35) means 30 training samples and 35 testing samples for feature extraction, UMI and classification method, HSN (UMI-HSN) for beta values from 0.30 to 0.75. The complete result can be showed in Figure 1.

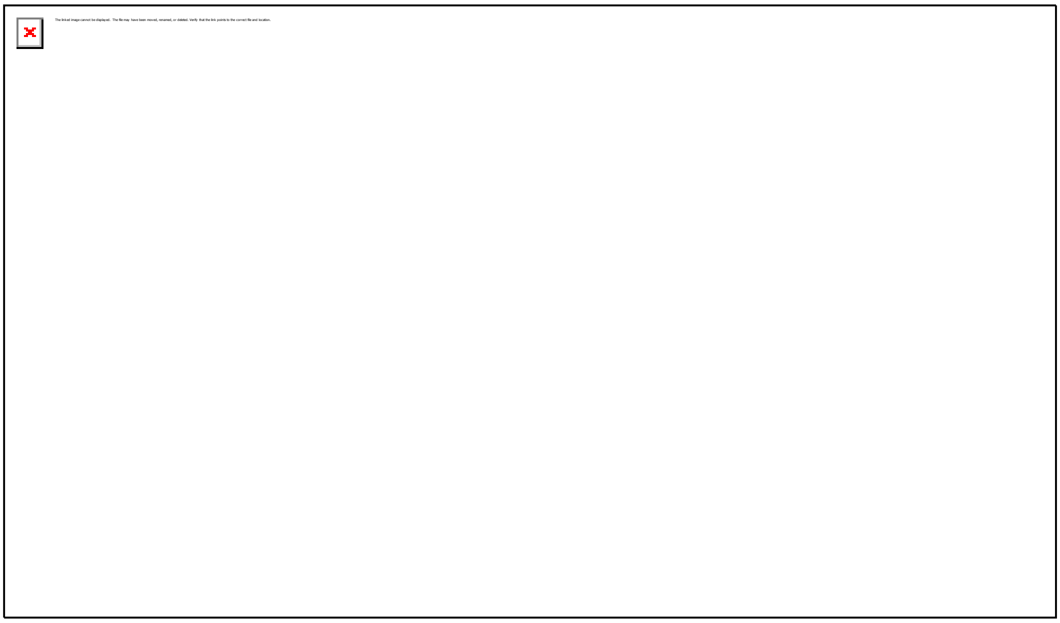

Figure 2. Bar Chart for UMI-HSN vs AUMI-HSN Based on Beta Values

Based on Figure 1, we make difference percentage correct rate between UMI-HSN method and AUMI-HSN method for beta values 0.75 as Table 3.

Table 3. The percentage matches the identification with UMI-HSN and AUMI-HSN

Data		Correct rate (%)	
Training Samples	Testing Samples	UMI-HSN	AUMI-HSN
20	25	88.00	96.40
30	25	86.00	97.20
20	35	89.71	95.14
30	35	88.00	97.14

Based on Table 3, correct rate UMI-HSN method for 25 testing samples with 20 and 30 training samples has the average result decrease from 88.00 to 86.00, and 35 testing samples with 20 and 30 training samples has the average result decrease from 89.71 to 88.00. This condition is different from AUMI-HSN method, the adding number of training samples can increase the percentage correct rate result.

CONCLUSION AND FUTURE WORK

This paper proposed AUMI-HSN and UMI-HSN for identifying the authorship of handwriting. The experiments in Table 3 showed that AUMI-HSN method was better than UMI-HSN method, the correct rate UMI-HSN was around 88% and AUMI-HSN was around 96%. Future work can be conducted to further explore the moment invariants feature extraction methods and cover neurons appropriate for BPR.

ACKNOWLEDGMENT

Authors would like to thank Research Management Centre (RMC) under vote RJ13000078284F026 and *Soft Computing Research Group (SCRG),* Universiti Teknologi Malaysia, for the support in making this study a success.

REFERENCES

Bulacu, M., & Schomaker, L. (2007). Text-Independent writer identification and verification using textural and allographic features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 4.

Jiang, J., Wei, H., & Qi, Q. (2009). Medical image segmentation based on biomimetic pattern recognition. *World Congress on Software Engineering*, 375-379.

Samsuryadi, & Shamsuddin, S.M. (2010). A framework of biomimetic pattern recognition in writer identification. *Proceedings of International Seminar of Information Technoloy, 35-41, Bandung, Indonesia.*

Shoujue, W. (2003). A new development on ANN in China – biomimetic pattern recognition and multi weight vector neurons. *Lecture Notes in Computer Science*, 35-43. Springer Verlag.

Shoujue, W., & Xingtao, Z. (2004). Biomimetic pattern recognition Theory and Its Applications. *Chinese Journal of Electronics*, 373-377, Vo1. 13, No. 3.

Xu, K., & Wu, Y. (2010). Motor imagery EEG recognition based on biomimetic pattern recognition. *Proceedings of BMEI 2010,3th International Conference on Biomedical Engineering and Informatics, 955-959.*

Yinan, S., Weijun, L., & Yuechao, W. (2003). United Moment Invariants for Shape Discrimination. *Proceedings of the 2003 IEEE, International Conference on Robotics, Intelligent Systems and Signal Processing*, 88-93.