

Bio-Inspired Generalized Global Shape Approach for Writer Identification

Azah Kamilah Muda, Siti Mariyam Shamsuddin, and Maslina Darus

Abstract—Writer identification is one of the areas in pattern recognition that attract many researchers to work in, particularly in forensic and biometric application, where the writing style can be used as biometric features for authenticating an identity. The challenging task in writer identification is the extraction of unique features, in which the individualistic of such handwriting styles can be adopted into bio-inspired generalized global shape for writer identification. In this paper, the feasibility of generalized global shape concept of complimentary binding in Artificial Immune System (AIS) for writer identification is explored. An experiment based on the proposed framework has been conducted to proof the validity and feasibility of the proposed approach for off-line writer identification.

Keywords—Writer identification, generalized global shape, individualistic, pattern recognition.

I. INTRODUCTION

IN the development of digital age, paper documents are still exchanged. In some situation, writer identification is needed to identify the original writer of handwritten document. Usually, writer identification performed on legal papers by a way of signature. However, it is also needed to identify handwritten documents without signature such as in threaten letter, authorship determination of old or historical manuscript etc. Handwriting has long been considered individualistic and writer individuality rests on the hypothesis that each individual has consistent handwriting [1]. Therefore, it ignited the researchers to explore this field in order identify the writer of handwriting. However, it still poses a challenge because human capability is superior when compared to computerized system in observing and recognizing the style of handwriting.

Meanwhile, Artificial Immune Systems (AIS) is one of the recent biologically inspired approaches that emerge from computer science field. AIS computational technique has led to the development of useful computational tools to solve complex problems such as in pattern recognition, fault detection, classifications, computer security, and optimization [2], [3], [4], [5]. This paper is intended to explore the feasibility of Negative Selection Algorithm and

Generalized concept in AIS for off-line handwritten writer identification domain.

This paper is organized into several sections. Writer identification is briefly described in Section 2, followed by AIS in pattern recognition in Section 3. Section 4 describes a proposed framework of bio-inspired writer identification and experiment conducted. And finally, conclusion is drawn in Section 5.

II. WRITER IDENTIFICATION

Writer identification is one of the areas that attract many researchers to work in. It is utilized in forensic and biometric application where an author of a questioned document can be identified based on handwriting. The identification of author for questioned handwritten document has a great importance on the criminal justice system and widely explored in forensic handwriting analysis [6], [7], [8], [9], [10]. The relation of character, shape and the styles of writing are different from one person to another. Features of handwriting are different according to these varieties of handwriting styles. However, handwriting is a skill that is personal to individual characteristics [11], [12] and it is individualistic [6], [9], [13], [14], [15]. It is a challenge to find the best solution in order to identify the writer accurately.

Most of the researchers in pattern recognition tried to solve writer identification problem based on the image processing and pattern recognition technique [1], [16], [17], [18], [19]. There is a close relationship between the tasks of writer identification and general handwriting recognition [16]. It involved typical pattern recognition framework, which are feature extraction and classification task [17], [18], [19]. In pattern recognition community, it is a well known that feature extraction and classification task are essential to achieve a good performance in recognizing patterns.

The main issue in writer identification is how to acquire features that reflect the author for these varieties of handwriting [9], [16], [17], [20], [21], [24], either for one writer or many writers. These features are required to classify in order to identify which group or classes that they are closed to. The shape or style of writing from one person to another is different. Even for one person, the styles are different in times. However, as mention before everyone has their own style of writing and it is individualistic. Each person's handwriting is seen as having a specific texture [15]. It must be a unique feature that can be generalized for each person in handwriting.

Previous works have been done in order to acquire the features from the handwriting such as, local graphemes

Manuscript received Ogos 31, 2006. This work is supported by STGL-012-2006, SAGA GRANT (Scientific Advancement Fund Allocation), Academy of Science, Malaysia.

Azah Kamilah Muda is with the University Technology of Malaysia, 81310 Skudai, Johor, Malaysia (e-mail: azah@kutkm.edu.my).

Siti Mariyam Shamsuddin is with the University Technology of Malaysia, 81310 Skudai, Johor, Malaysia (e-mail: mariyam@fsksm.utm.my).

Maslina Darus is with the Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia (e-mail: maslina@pkrisc.cc.ukm.my).

features [20]; text line geometrical quantity [16]; dynamic features [24]; character formulated [8]; texture analysis [21], [17]; text line based features [23], [14]; but none of them focus on generalized global shape features from handwriting.

III. AIS IN PATTERN RECOGNITION

AIS draw inspiration from vertebrate immune system, the same as artificial neural network from nervous system. (de Castro and Timmis, 2002b) defined AIS as an adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving. The natural immune system is an adaptive learning system that employs many parallel and complementary mechanisms for defense against foreign pathogens. It is a distributed system, capable of learning to identify previously unseen invaders and remembering what it has learnt. Numerous immune algorithms now exist, based on processes identified within human immune systems.

In the work of pattern recognition, immune recognition is based on the complementary bind region between the receptor and a portion of the antigen called epitope. It is important to note that the recognition in the immune system is based on shape complementary [2]. This recognition concept in Immune System (IS) is performed by using 3 mechanisms which are (i) Negative Selection (T-cells that recognize self-antigens are excluded from the population of T-cells during the maturation process); (ii) Clonal Selection (if B-cell encounters a non-self antigen with a sufficient affinity, it will proliferates and differentiates into memory cells) and (iii) Immune Network (if B-cell recognizes a self-antigen, it might result in suppression). However, this paper will only focus on NSA. NSA in AIS has been applied to detect computer viruses [7]; tool breakage detection and time-series anomaly detection [9]; network intrusion detection [10], [11]; color image classification [12], creative design classification [13], cancer classification [22].

A. Negative Selection Algorithm

Negative Selection Algorithm was proposed by Forrest in [25] for various anomaly detection problems. It was inspired by Negative Selection property in human immune system that uses the property of self/non-self discrimination to detect foreign antigens. It presents an alternative paradigm to perform pattern recognition by storing information about the complement set (non-self cell) of the pattern to be recognized. It also provides tolerance for self cells and deal with the immune system's ability to detect unknown antigens while not reacting to the self cells. In the biological system, this is achieved in part by T-cells which have receptors on their surface that can detect foreign antigens.

Negative Selection Algorithm works on similar principles, generating detectors randomly and eliminating the ones that detect self, so that the remaining non-self cells can detect any non-self. It has two main stages which are censoring and monitoring. The first stage of NSA is to control the generation of detectors in random and the other one is to monitor the changes by using the detectors generated in earlier stage. The matching of detector set with new antigens based on certain matching rule. Non-self is detected if there is a match between the antigen and any of

the detectors. The NSA by [25] and summarized by [27] is as follows:

- 1) Define self as a collection of strings S of finite length L that needs to be monitored.
- 2) Generate a set of detectors R each of which fails to match any self string in S .
- 3) Monitor S for changes by continually matching the detectors in R against S . If any detector matches, then a change is known to have occurred, because the detectors are designed not to match any of the original strings in S .

According to Xiao in [26], the original NSA is not suitable for pattern recognition task. The NSA is originally used to solve change detection problems with only two states defined, either normal or abnormal. In pattern recognition task, negative selection must be extended to multi-state, equivalent to number of classified groups. Xiao in [26] modified the original NSA to be used for pattern recognition in engineering creative design.

B. Generalization in AIS

Generalized shape concept is important in immune system where an antibody can recognize an antigen based on the shape. Antigens and antibodies are considered to be points in an abstract shape space, where coordinates of points in the space represent generalized physico-chemical properties associated with various physical properties related to binding, such as geometric shape [27]. The affinity between an antibody and an antigen involve binding process where it must be bind complementary with each other. However, it must not be completely bind between these two elements, to recognize the antigen or virus. Antibody can detect the virus via regions of complementary with a set of features that called generalized shape of a molecule [28]. Fig. 1 below illustrates the generalized shape concept of complementary binding in AIS. This generalization capability can recognize not only the specific pattern, but also any structurally related pattern. This capability is attained by a process called cross-reactivity and can be modeled using the affinity threshold [29].

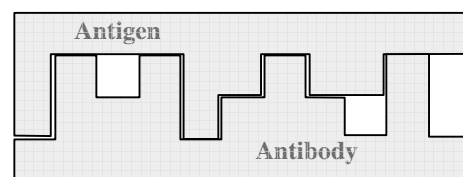


Fig. 1 Generalized shape concept of complementary binding in AIS

Generalization can be seen as the capability of a input data set respond appropriately to another set of data. A given attribute of string data can match not only with the exact or complete complementary, but also to an acceptable complementary process within the given threshold. The volume resulting from the cross-reactivity threshold is called recognition region. If the value of affinity binding for recognition region is larger than or equal to the pre-define cross-reactivity threshold, then it is assumed that a recognition event is occurred between the molecules in the

system. In computational system, each antibody and antigen is assumed to be implicitly described by a vector of numbers i.e. a coordinate vector, which represent the geometric shape characteristics relevant to shape complementary in binding [27].

IV. PROPOSED BIO-INSPIRED FRAMEWORK AND EXPERIMENT

A simple experiment was conducted base on the proposed framework in Fig. 2. This experiment is to proof the feasibility of bio-inspired generalized approach in writer identification. It has three main tasks, which are Granular Data Collector, Bio-inspired Training Environment and Bio-inspired Classifier.

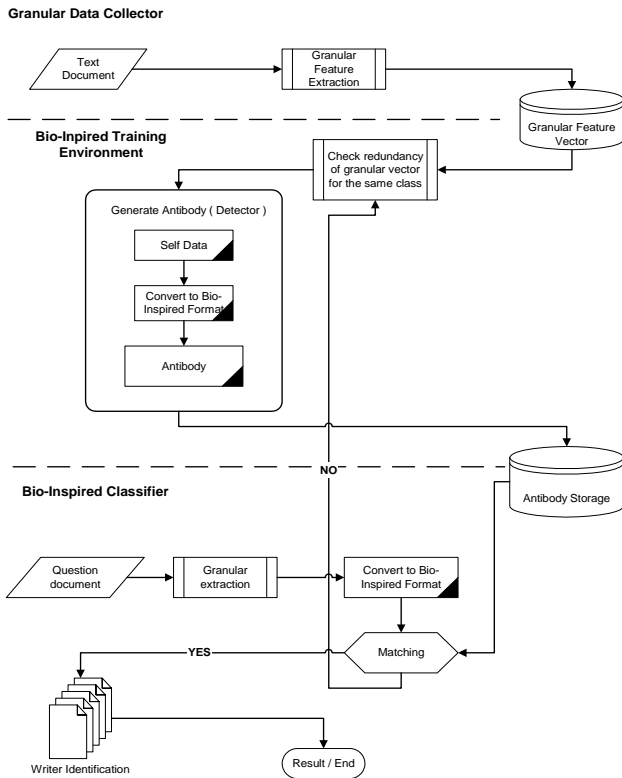


Fig. 2 Proposed Framework of Bio-inspired Writer Identification

A. Granular Data Collector Phase

In the task Granular Data Collector, global moments function specifically on Geometric Moment Function has been used to extract the handwriting image. A set of moments computed from digital image generally represents global characteristics of the image shape, and provides a lot of information about different types of geometrical features of the image [30]. The invariants feature vectors are converted using the mathematical formulation of $f: \mathbb{R} \rightarrow \sum^m$ into symbolic representation.

Image of words need to be transformed into suitable format in order to be manipulated and processed. Feature extraction is performed on those images using moment functions to generate respective invariant values. Table I

shows an example of word 'since' representation with Geometric Moment Invariant (GMI) technique. These features values were the results of feature extraction of the images using respective moment functions.

TABLE I
DATA REPRESENTATION OF 'SINCE' USING GMI

Feature Vector	
1	-1.780028
2	-9.653448
3	-14.781119
4	-15.007411
5	-30.545805
6	0.000000
7	-1.780028

B. Bio-Inspired Training Phase

In Bio-inspired Training phase, sets of self data are obtained from first task. These data sets are transformed and represented into Bio-Inspired Format by performing the modified NSA to generate antibody data set. The modified NSA used in this work is shown as below:

- Define self as N sets of strings S is used for pattern matching (how many classes according to how many writer is observed).
- Generate N sets of detectors in complimentary and each should fail to match with self set define in [i].
- Present the input pattern to monitor process - if any match occurs at the i^{th} set of detectors, the input pattern belongs to the corresponding groups. The process is repeated until the last writer in handwriting database.

The identified antibodies are generated and stored for matching procedure with antigen in the classification task.

C. Classification Phase

In classification phase, the concept of generalized global shape will be applied where the shape of handwritten word is recognized with the complementary binding of self cell and antibody. The questioned document is extracted, and the features are transformed into binary format as antigens. This will be defined and act as antigens (or as self data for the examined word image). A binary matching rule is defined in terms of individual bit matching of detectors and antigens represented as binary string [31].

In the process, r-Contiguous matching technique is used. The binary matching of r-contiguous is defined as: Given

$$\text{Antigen } x = x_1 x_2 x_3 \dots x_n$$

and

$$\text{Detector } d = d_1 d_2 d_3 \dots d_n,$$

where,

$$d \text{ matches } x \equiv \exists i \leq n - r + 1$$

such that

$$x_j = d_j \text{ for } j = i, \dots, i + r - 1$$

The two strings are match if there is a sequence of size 2, where all the bits are identical. For negative selection using XOR operator, matching is achieved with complements, i.e., 0 matches with 1. These antigens are compared to antibody from the antibody database. If there is no matched, then the antigen will go through the generate antibody process to produce a new data (memory). The new data are saved to the database. Otherwise, it will be listed as potential writer which will later be sorted by percentage. The highest percentage is claimed as the potential owner of the questioned document.

In this initial experiment, 10 writers with 20 words of each person (all 200 words) from AIM database has been used in training data set. All the extracted features from these writers are saved in granular feature vector database. For data testing, the different set of 5 words image from writer no 10 is extracted to be used as features from question document. Result from the experiment is shown in Table II. The accuracy of word 1 to word 4 for the 10th writer is the highest in accuracy for all writers, accept for word 5. However, the differences of accuracy between writer no 9 and writer no 10 is only 5%. In overall, it shows that the writing from question document can identify the 10th writer as the owner of the writing. Thus, it is proven the feasibility of deploying the bio-inspired generalized global shape approach in writer identification.

TABLE II
RESULT FROM THE EXPERIMENT

Writer	Words Image Accuracy in %				
	word 1	word2	word3	word4	word5
W1	65	71	58	51	63
W2	72	67	67	45	51
W3	68	70	73	59	58
W4	70	70	59	48	56
W5	66	67	68	54	48
W6	65	70	69	58	57
W7	67	67	60	51	55
W8	65	70	60	47	54
W9	77	72	66	62	69
W10	98	98	78	65	64
Target Writer	W10	W10	W10	W10	W10
Result Writer	W10	W10	W10	W10	W9

V. CONCLUSION AND FUTURE WORK

We have presented a bio-inspired of generalized shape with complimentary binding concept in AIS for identifying the potential writer of handwriting. GMI is utilized as the feature extraction technique to extract global shape features of handwritten word. An initial experiment has been conducted to explore and proof the feasibility of proposed approach. In future work, we will use other feature extraction techniques and a bigger set of data to perform the experiment for comparison.

REFERENCES

[1] Srihari, S.N.; Sung-Hyuk Cha and Sangjik Lee; (2001). "Establishing handwriting individuality using pattern recognition techniques",

Vol. 2, No. 4, 2001. Analysis and Recognition, 2001. Proceedings. Sixth International Conference on, 10-13 Sept. 2001, Pages: 1195 – 1204.

[2] de Castro, L.N. and Timmis, J.I. "Artificial immune systems as a novel soft computing paradigm". *Soft Computing Journal*. 2003. 7: 526-544.

[3] de Castro, L. N. and Timmis, J. I. "Artificial Immune Systems: A Novel Paradigm to Pattern Recognition", *Artificial Neural Networks in Pattern Recognition*, J. M. Corchado, L. Alonso, and C. Fyfe (eds.), SOCO-2002, University of Paisley, UK, pp. 67-84.

[4] Olfa Nasaroui, Fabio Gonzalez, and Dipankar Dasgupta. "The Fuzzy Artificial Immune System: Motivations, Basic Concepts, and Application to Clustering and Web Profiling". *IEEE International Conference on Fuzzy Systems. IEEE World Congress on Computational Intelligence, Hawaii*. 2002. pp. 12-17.

[5] de Castro, L. N. and Von Zuben, F. J. "Artificial Immune Systems: Part II – A Survey of Applications", *Technical Report – RT DCA 02/00*, 2000. pp. 65.

[6] Srihari, S. N.; Huang, C.; Srinivasan H. and Shah, V. A. (2006). "Biometric and Forensic Aspects of Digital Document Processing", *Digital Document Processing*, B. B. Chaudhuri (ed.), Springer.

[7] Pervouchine, V., Leedham, G and Melikhov, K. (2005). "Handwritten character skeletonisation for forensic document analysis", *Proceedings of the 2005 ACM symposium on Applied computing*, March 13–17, 2005, Santa Fe, New Mexico, USA.

[8] Tapiador, Marino and Sigüenza, Juan A. (2004). "Writer Identification Method Based on Forensic Knowledge". *Biometric Authentication: First International Conference, ICBA 2004*, Hong Kong, China, July 15-17, 2004. *Proceedings* pp. 555 – 561.

[9] Srihari, S. N.; Cha, S.-H.; Arora, H.; and Lee, S.; (2002). "Individuality of Handwriting", *Journal of Forensic Sciences*, 47(4), July 2002, pp. 1-17.

[10] Franke, Katrin and Köppen, Mario, (2001). "A computer-based system to support forensic studies on handwritten documents", *International Journal on Document Analysis and Recognition*, Volume 3, Number 4, Pages: 218 – 231.

[11] Rohlik, Ondrej. "Handwritten Text Analysis.", University of West Bohemia, Pilsen. : Ph.d Dissertation. 2003.

[12] Plamondon, Rejean.; Srihari, Sargur. N.; (2000). "On-line and off-line handwriting recognition: A comprehensive survey", *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on* Volume: 22, Issue: 1, Jan. 2000, Pages:63 – 84.

[13] Bin, Zhang and Srihari, S. N., (2003). "Analysis of Handwriting Individuality Using Word Features". *Document Analysis and Recognition*, 2003. *Proceedings. Seventh International Conference on* 3-6 Aug. 2003 Page(s):1142 - 1146.

[14] Marti, U.-V.; Messerli, R.; Bunke, H.; (2001). "Writer identification using text line based features", *Document Analysis and Recognition*, 2001. *Proceedings. Sixth International Conference on* 10-13 Sept. 2001 Page(s):101 – 105.

[15] Yong Zhu; Tieniu Tan; Yunhong Wang; (2000). "Biometric Personal Identification Based on Handwriting", *Pattern Recognition*, 2000. *Proceedings. 15th International Conference on* Volume 2, 3-7 Sept 2000 Page(s):797 - 800 vol.2.

[16] Schlappbach, Andreas and Bunke, Horst. (2004b). "Off-line Handwriting Identification Using HMM Based Recognizers", In *Proc. 17th Int. Conf. on Pattern Recognition*, 23-26, August, 2004. pp. 654–658, Cambridge, UK.

[17] Shen, Cong; Ruan, Xiao-Gang and Mao, Tian-Lu. (2002). "Writer identification using Gabor wavelet", *Intelligent Control and Automation*, 2002. *Proceedings of the 4th World Congress on* Volume 3, 10-14 June 2002 Page(s):2061 - 2064 vol.3.

[18] Bensefia, A.; Paquet, T. and Heutte, L. (2005a). "A writer identification and verification system", *Pattern Recognition Letters*, In Press, Corrected Proof, Available online 23 May 2005.

[19] Said, H.E.S.; Tan, T.N. and Baker, K.D. "Writer identification based on handwriting", *Pattern Recognition* 33, 2000, pp.149-160.

[20] Bensefia, A.; Nosary, A.; Paquet, T and Heutte, L., (2002). "Writer identification by writer's invariants", *Frontiers in Handwriting Recognition*, 2002. *Proceedings. Eighth International Workshop on* 6-8 Aug. 2002 Page(s):274 – 279.

[21] He, Z.Y.; and Tang, Y.Y. (2004). "Chinese handwriting-based writer identification by texture analysis". *Machine Learning and Cybernetics*, 2004. *Proceedings of 2004 International Conference on* Volume 6, 26-29 Aug. 2004 Page(s):3488 - 3491 vol.6.

[22] Ando, Shin and Iba, Hitoshi. (2003). "Artificial Immune System for Classification of Cancer", *EvoWorkshops 2003*. LNCS 2611, pp. 1–10.

- [23] Wirotius, M.; Seropian, A.; Vincent, N., (2003). "Writer identification from gray level distribution", Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on 3-6 Aug. 2003 Page(s):1168 – 1172.
- [24] Kun Yu; Yunhong Wang; Tieniu Tan; (2004). "Writer Identification Using Dynamic Features", Biometric Authentication: First International Conference, ICBA 2004, Hong Kong, China, July 15-17, 2004. Proceedings. pp. 512 – 518.
- [25] Forrest, S., Perelson, A.S., Allen, L. and Cherukuri, R. "Self-nonsel self discrimination in a computer". Proceedings of IEEE Symposium on Research in Security and Privacy, 16-18. 1994 May pp. 202-212. Oakland, CA USA: IEEE.
- [26] Xiao, R. B., Wang, L., and Liu, Y. "A Framework of AIS Based Pattern Classification and Matching for Engineering Creative Design". Proceedings of the First International Conference on Machine Learning and Cybernetics. Beijing: IEEE. 2002. pp. 1554-1558.
- [27] Dasgupta, D. "Artificial Immune Systems and Their Applications". Germany: Springer-Verlag. 1999.
- [28] de Castro, L. N., & Timmis, J. (2002b), Artificial Immune Systems: A New Computational Intelligence Approach, Springer-Verlag.
- [29] de Castro, L. N. & Timmis, J. I. (2002a). "Artificial Immune Systems: A Novel Paradigm to Pattern Recognition", Artificial Neural Networks in Pattern Recognition , J. M. Corchado, L. Alonso, and C. Fyfe (eds.), SOCO-2002, University of Paisley, UK, pp. 67-84.
- [30] Mukundan, R and Ramakrishnan, K.R. (1998). "Moment Functions in Image Analysis Theory and Application". Singapore: World Scientific Publishing Co.Pte.Ltd.
- [31] González, F., Dasgupta, D. and Gómez, J. (2003). The Effect of Binary Matching Rules in Negative Selection. GECCO, LNCS 2723. 195-206.

Azah Kamilah received the BSc and MSc from Universiti Teknologi Malaysia (UTM), Malaysia in 1997 and 1999 respectively. From 1997 to 2002 she joined the Department of Software Engineering at Faculty of Computer Science and Information System, Universiti Teknologi Malaysia as a lecturer. Since 2002 he has been with Kolej Universiti Teknikal Kebangsaan Malaysia as a lecturer in Faculty of Information Technology and Communication. Now she is doing her PhD in bio-inspired pattern recognition of writer identification at Universiti Teknologi Malaysia. Her research interests are Pattern Recognition, Soft & Biologically Inspired Computing, System Identification and Artificial Intelligence.

Associate Prof. Dr. Siti Mariyam Shamsuddin received her Bachelor and Master degree in Mathematics from New Jersey USA, and Phd in Pattern Recognition & Artificial Intelligence from Universiti Putra Malaysia (UPM), Malaysia. Currently, she is a Head of ICT Cluster, Universiti Teknologi Malaysia (UTM). Her research interests include Pattern Recognition, Soft & Biologically Inspired Computing and Geometric Modeling.

Associate Prof. Dr. Maslina Darus received her Bachelor degree in Mathematics from Acadia University, Canada, and Phd in Pure Mathematics (Complex Analysis) from University of Wales, Swansea, U.K. Currently, she is Head of Mathematics Programme, School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia (UKM). Her research interests include Geometric Function Theory and Geometric Moment Theory.