An Approach of Bio-inspired Generalized Shape for Writer Identification

Azah Kamilah Muda & Siti Mariyam Shamsuddin
Faculty of Computer Science & Information System,
University Technology of Malaysia, 81310 Skudai, Johor, Malaysia.
azah@kutkm.edu.my; mariyam@fsksm.utm.my

Abstract
Writer identification is one of the areas that attract many researchers to work in. It has being focus in forensic and biometric application where the writing styles can be used as biometric features for authenticating an identity. The individualistic of handwriting styles can be adopted into bio-inspired generalized global shape in writer identification. This paper is discussed the problem overview and solution concept of generalized global shape in writer identification. Moreover, to explore the feasibility of this generalized global shape approach in this domain. An experiment based on the proposed framework has been conducted to proof the feasibility of this approach in off-line writer identification.

Keyword: writer identification, generalized global shape approach, individualistic, biometric application

1.0 Introduction

In the development of digital age, the paper documents are still exchanged. In some situation, writer identification is needed to identify the original writer of a handwritten document. Usually, writer identification performed on legal papers by a way of signature. However, it is also a needed to identify a handwritten document without signature such as in threaten letter, authorship determination of old or historical manuscript etc. Handwritten identification can be included as a particular kind of dynamic biometric where the shapes and the writing styles can be used as biometric features for authenticating an identity [1], [2], [3], [4]. Handwriting has long been considered individualistic and writer individuality rests on the hypothesis that each individual has consistent handwriting [5]. It ignited the researchers to explore this field in order identify the writer of handwriting. It still poses a challenge because human capability is beyond to computerized system in observing and recognizing the style of handwriting.

Meanwhile, Artificial Immune Systems (AIS) is one of the recent biologically inspired approaches to emerge from computer science. 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. AIS computational technique has led to the development of useful computational tools for the solution of complex problems such as in pattern recognition, fault detection, classifications, computer security, and optimization [6], [7], [8], [9].

In the work of pattern recognition, Immune System (IS) is performed by using 3 mechanisms which are Negative Selection (T-cells that recognize self-antigens are excluded from the population of T-cells during the maturation process), Clonal Selection (if B-cell encounters a non-self antigen with a sufficient affinity, it will proliferates and differentiates into memory cells) and Immune Network (if B-cell recognizes a self-antigen, it might result in suppression). This paper is to explore the feasibility of generalized concept in AIS into off-line handwritten writer identification domain.

This paper is organized into several sections. Writer identification is briefly described in Section 2, followed by basic immune system recognition in Section 3. Section 4 discussed in brief about AIS in pattern recognition. Section 5 described a proposed framework of bio-inspired writer identification and experiment conducted. Finally, conclusion is drawn in Section 6.

2.0 Writer Identification

Writer identification is one of the areas that attract many researchers to work in. The relation of character, shape and the styles of writing are different from one to another. Features of handwriting are different according to these varieties of handwriting styles. It makes it a challenge to find the best solution in order to identify the writer accurately. However, handwriting is a skill that is personal to individual characteristics [13], [14] and it is individualistic [1], [15], [11], [16], [4]. Most of the researchers in pattern recognition tried to solve writer identification problem based on the image processing and pattern recognition technique [17], [18], [19], [5], [20]. There is a close relationship between the tasks of writer identification and general handwriting

recognition [17]. [18], [19] and [20] mentioned it involve typical pattern recognition framework, which are feature extraction and classification task. In pattern recognition community, it is a well known that feature extraction and classification task are important to achieve a good performance in recognizing patterns.

3.0 Trends Leading to the Problem

The shape or style of writing from one person to another is different. Even for one person, they 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 [4]. It must be a unique feature that can be generalized for each person in handwriting, even though one person has many styles of writing at different times as well. The main issue in writer identification is how to acquire the features that reflect the author for these varieties of handwriting [21], [17], [3], [22], [11], [18]; 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.

Many previous work have been done in order to acquire the features from the handwriting where they focus on rigid characteristics such as local graphemes features [21]; text line geometrical quantity [17]; dynamic features [3]; character formulated [2]; texture analysis [22], [18]; text line based features [23], [16], but none of them focus on generalized global features of word shape as one whole from writing. These rigid characteristics are contributed to the large lexicon. A common behavior of the actual systems is that the accuracy decreases as the number of reference vector in lexicon grows. Furthermore, the computational complexity is also related to the lexicon, and it increases relatively to its size [35]. Global approach will not incur additional lexicon into the database [36]; [37]; [38]. The background and trends leading to the problem in this study, is shown in Figure 1.

4.0 AIS in Pattern Recognition

As mentioned before, IS has three mechanisms in recognizing pattern but, this study will only focus on NSA. NSA in AIS has been applied to detect computer viruses [10]; tool breakage detection and time-series anomaly detection [11;] network intrusion detection [12], [13]; color image classification [14] and creative design classification [15]. The original work by [10] which negative selection algorithm was proposed has been inspirational to almost all the research in the AIS related to computer security. In pattern recognition, most of the researches are also influenced by NSA besides Clonal Selection Algorithm (CSA).

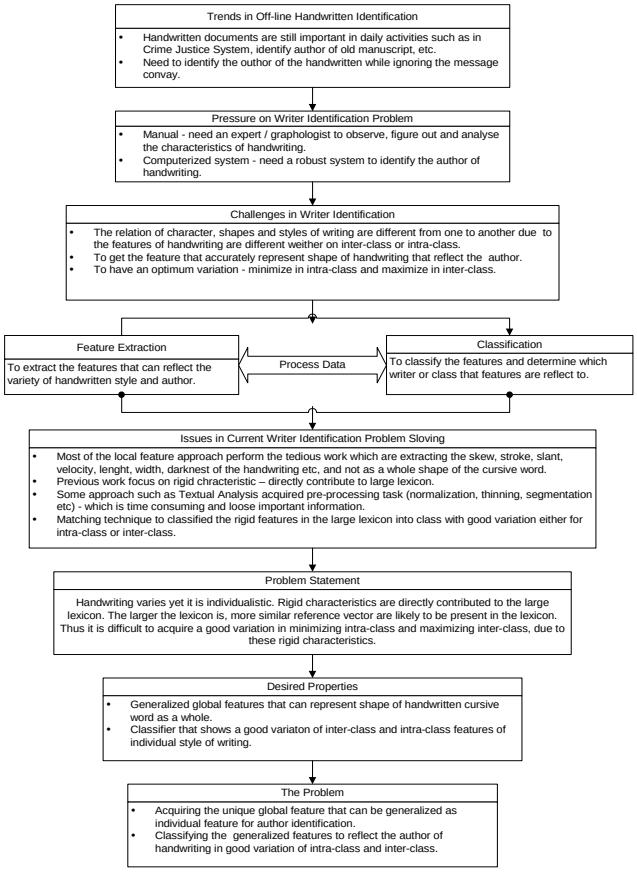

Figure 1 : Background that Leads to the problem

4.1 Negative Selection Algorithm

NSA was inspired by Negative Selection property in human immune system where 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 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. A pseudo random genetic process makes the receptors during the generation of the T-cells [31]. Then they undergo a process of filtering in the thymus where T cells that react against self cells are destroyed and only those do not bind to self cells are allowed to leave thymus. These matured T cells circulate through out the body to protect against 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 proposed by [25] is originally used to solve change detection problems with only two states defined, either normal or abnormal. However, according to Xiao in [30], the original NSA is not suitable for pattern recognition task. Xiao modified the original NSA to be used for pattern recognition in engineering creative design. It must be extended to multi-state, equivalent to number of classified groups. In this writer identification study, no random detectors are generated and compared with self string to determine the antibody. This is because simply by keeping the complement of self string as antibody is enough to cover all detectors possibly generated using random method. The antibody sets are much smaller compared to random method. For instance, if there are 10 elements in self set, then only 10 antibodies are generated with complement. Random detectors generation has the limitation of generating valid detectors if the size of random generated detectors is small [29]. However, if large number of detectors is generated, it will be time consuming and there is no guarantee that generated detectors are not repeated in the process. Other work on classification task that relate to this pattern recognition that used NS, still perform randomly generated such as in image color classification [29] and cancer classification [39]. Table 1 below shows the comparison of original NSA by [25] for anomaly detection problems and Modified NSA by [30] for engineering creative design problem and the proposed Modified Immune Classifier WI problem in this study.

Table 1 : Comparison of Negative Selection Algorithm

Researcher	Algorithm	Note
Traditional NSA (Forrest et. al, 1994)	i. Define self set string → two class only. ii. Generate a set of detectors in random that fails to match with any self string in [i]. iii. Monitor with matching process - if any detector matches, then recognition process occurred in detecting antigen / non-self , because the detectors are designed not to match any of the original strings.	Originally to solve anomaly detection problem thus, self set strings are defined only two types of states which are normal and abnormal. Randomly generate detectors.
Modified NSA (Xiao, 2002)	i. Define self as N sets of strings is used for pattern matching → how many classes according to how many class involved. ii. Randomly generate N sets of detectors and each should fail to match with members of the pre-classified groups. iii. 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 and the pattern matching task is finished.	Number of self set string is defined according to how many classes of object in that domain problem. Randomly generate detectors. Monitor process stop immediately after the recognition occurred
Modified Immune Classifier	i. Define self as N sets of strings is used for pattern matching → how many classes according to how many writer is observed. ii. Generate N sets of detectors in complimentary and each should fail to match with self set define in [i]. iii. 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 until the last writer in handwriting database.	Number of self set string is defined according to how writer will be observed in database. Complimentary process in generating detectors. Classifying or detecting process will be run until the last data to get the highest accuracy → will be the potential writer that matches with the writer of question document.

4.2 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 this space represent generalized physico-chemical properties associated with various physical properties related to binding, such as geometric shape [32]. The affinity between an antibody and an antigen involve binding process where it must 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 [33]. Figure 2 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 [34].

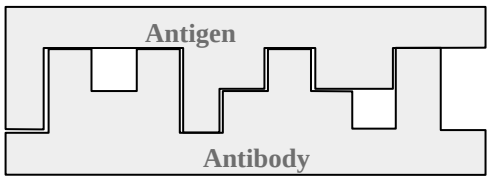

Figure 2 : Generalized shape concept of complementary binding in AIS

Generalization can be seen as the capability of a set of input data respond appropriately to another set of data. A given attribute string of data can match not only the exact complement but also the reasonable complementary to it. 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 cross-reactivity threshold that defined earlier, 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 [32].

5.0 Proposed Bio-inspired Framework and Experiment

An initial experiment was conducted based on the proposed framework, as shown in Figure 3. This experiment is to show feasibility of bio-inspired generalized approach in writer identification.

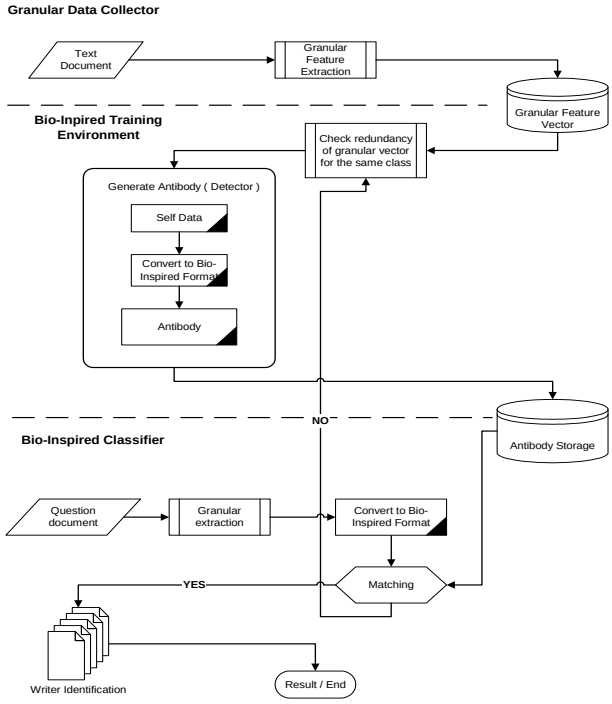

Figure 3 : Proposed Framework of Bio-inspired Writer Identification

In the experiment, geometrical moment function is performed in feature extraction task, meanwhile in identification or classification task, r-Contiguous matching technique is used. 10 writers with 20 words of each person 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 nombor 10 is extracted to be used as features from question document. Result from the experiment is shown in Table 2.

The accuracy of word 1 to word 4 for the 10^{th} writer has highest accuracy, 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 10^{th} writer as the owner of the writing. Thus, it is proven that the feasibility of deploying the bio-inspired generalized global shape concept in writer identification is match.

Table 2 : Result from the Experiment

Writer	Word Image Accuracy in %				
	Word 1	Word 2	Word 3	Word 4	Word 5
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

6.0 Conclusion and future work

We have presented an approach of bio-inspired generalized global shape in identifying the potential writer of handwriting. An initial experiment has been conducted to explore and proof the feasibility of proposed approach with geometrical function. In future work, we will try to use the other feature extraction technique and larger data to perform the experiment in this study.

7.0 References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] Srihari, S.N.; Sung-Hyuk Cha and Sangjik Lee; (2001). "Establishing handwriting individuality using pattern recognition techniques", Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on , 10-13 Sept. 2001, Pages:1195 – 1204.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] Schlapbach, 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

[18] 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.

[19] 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.

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

[21] 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.

[22] 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.

[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] Nossal, G. J. V. "Life, Death and the Immune System", Scientific American, 1993. 269(3), pp. 21-30.

[25] Forrest, S., Perelson, A.S., Allen, L. and Cherukuri, R. "Self-nonself 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] Dasgupta, D. "An Overview of Artificial Immune Systems and Their Applications", Berlin: Springer-Verlag. 3-23. 1998

[27] Hofmeyr, S., (1999). "An Immunological Model of Distributed Detection and Its Application to Computer Security". University of New Mexico: Phd Thesis.

[28] Hofmeyr, S., and Forrest, S. "Architecture for an Artificial Immune System". Evolutionary Computation, 2000. **7**(1). 45-68.

[29] Sathyanath,S., Sahin, F. "An AIS Approach to a Color Image Classification Problem in a Real Time Industrial Application". IEEE International Conference on Systems, Man, and Cybernetics. 4. 7-10. 2001. pp.2285-2290. Tucson, AZ USA: IEEE.

[30] 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.

[31] Lapedes, Alan & Farber, Robert, (2001). "The Geometry of Shape Space: Application to Influenza", Journal of Theoretical Biology, Volume 212, Issue 1, 7 September 2001, Pages 57-69.

[32] Dasgupta, D. "Artificial Immune Systems and Their Applications". Germany: Springer-Verlag. 1999.

[33] de Castro, L. N., & Timmis, J. (2002b), Artificial Immune Systems: A New Computational Intelligence Approach, Springer-Verlag.

[34] 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.