UNIVERSITI PUTRA MALAYSIA

AUTOMATED PLANT RECOGNITION SYSTEM BASED ON MULTI-OBJECTIVE PARALLEL GENETIC ALGORITHM AND NEURAL NETWORK

SEYED MOHAMMAD HOSSEIN SEFIDGAR

FK 2014 67

AUTOMATED PLANT RECOGNITION SYSTEM BASED ON MULTI-OBJECTIVE PARALLEL GENETIC ALGORITHM AND NEURAL NETWORK

By

SEYED MOHAMMAD HOSSEIN SEFIDGAR

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

February 2014

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

AUTOMATED PLANT RECOGNITION SYSTEM BASED ON MULTI-OBJECTIVE PARALLEL GENETIC ALGORITHM AND NEURAL NETWORK

By

SEYED MOHAMMAD HOSSEIN SEFIDGAR

February 2014

Chair: Siti Anom Ahmad, PhD
Faculty: Engineering

Plant recognition system is a system that recognizes the species of plants automatically. The applications of this system are in medicine, botanical research and agriculture. In the recent years, lack of sufficient botanist increases the need for computerized system. Also, it can be seen that working with these systems are more convenient and quick when dealing with huge data. The problem with the existing plant recognition system is the lack of method to find the best structure for their classifiers. This work presents some contributions to plant recognition system. Number of samples involving Flavia, Citrus and Coleus were collected. Then, suitable features including texture and shape were extracted from the dataset. Texture features involved the middle energy and the middle entropy and shape features involved statistical characterizations including variance, median, standard deviation and mean. Next, the classification was carried out. First, the best set of structures for feed forward neural network were found by multi objective parallel genetic algorithm. This approach regarded three criteria involving mean square error, Akaike information criterion and minimum description length to rate different feed forward neural network structures and to select the best set of them. Lastly, feed forward neural network with the best structures were applied to classify the dataset. This method resulted around 99% of classification rate. To conclude, multi objective parallel genetic algorithm can automatically tune feed forward neural network to classify the dataset with a good classification rate.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

SISTEM PENGECAMAN TUMBUHAN AUTOMATIK BERDASARKAN ALGORITMA GENETIK SELARI PELBAGAI-OBJEKTIF DAN RANGKAIAN NEURAL

Oleh

SEYED MOHAMMAD HOSSEIN SEFIDGAR

Februari 2014

Pengerusi: Siti Anom Ahmad, PhD
Fakulti: Kejuruteraan

Sistem pengecaman tumbuhan adalah satu sistem yang mengecam spesies tumbuhan secara automatik. Aplikasi sistem ini adalah dalam bidang perubatan, penyelidikan botani dan pertanian. Sejak kebelakangan ini, kekurangan ahli botani meningkatkan keperluan sistem berkomputer. Selain itu, ia boleh dilihat bahawa bekerja dengan sistem ini adalah lebih mudah dan cepat apabila berurusan dengan data yang besar. Masalah dengan sistem pengiktirafan tumbuhan adalah kekurangan kaedah untuk mencari struktur yang terbaik untuk penjodoh bilangan mereka. Kerja ini membentangkan beberapa sumbangan untuk menanam sistem pengiktirafan. Bilangan sampel yang melibatkan Flavia, Citrus dan Coleus telah dikumpulkan. Kemudian, ciri-ciri yang sesuai termasuk tekstur dan bentuk ini diperolehi daripada dataset itu. Ciri-ciri tekstur yang terlibat tenaga pertengahan dan entropi pertengahan dan bentuk ciri-ciri yang terlibat pencirian statistik termasuk varians, sisihan piawai dan min. Seterusnya, pengelasan yang telah dilakukan. Pertama, set terbaik struktur untuk makanan rangkaian neural ke hadapan ditemui oleh objektif algoritma selari genetik berbilang. Pendekatan ini dianggap tiga kriteria yang melibatkan ralat min persegi, maklumat kriteria Akaike dan panjang deskriptif minimum untuk mengadar makanan ke hadapan struktur rangkaian neural yang berbeza dan untuk memilih set yang terbaik daripada mereka. Akhir sekali, makanan rangkaian neural ke hadapan dengan struktur yang terbaik telah digunakan untuk mengelaskan dataset itu. Kaedah ini menyebabkan sekitar 99 % daripada kadar pengelasan. Untuk menyimpulkan, objektif algoritma selari genetik berbilang boleh secara automatik lagu makanan rangkaian neural ke hadapan untuk mengklasifikasikan dataset dengan kadar pengelasan yang baik.

ACKNOWLEDGEMENTS

To begin with, I would like thank my supervisor Dr. Siti Anom Ahmad for all her support throughout my project.

Words fail me to express my appreciation to my parents for their inseparable support, prayer and my beloved sister Sara.

My special thanks goes to all my friends. Finally, I would like to thank all my colleagues for their valuable assistance in my project.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science.
The members of the Supervisory Committee were as follows:

Siti Anom Binti Ahmad, PhD
Senior Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Chairperson)

Raja Mohd Kamil Bin Raja Ahmad, PhD
Senior Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Member)

____

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

Declaration by the student

I hereby confirm that:

- this thesis is my original work
- quotations, illustrations and citations have been duly referenced
- the thesis has not been submitted previously or comcurrently for any other degree at any institutions
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be owned from supervisor and deputy vice –chancellor (Research and innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature: _______________________ Date: __________________

Name and Matric No.: ____

Declaration by Members of Supervisory Committee

This is to confirm that:

- The research conducted and the writing of this thesis was under our supervision;
- Supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ____________________	Signature: ____________________
Name of Chairman of Supervisory Committee: **Siti Anom Ahmad**	Name of Member of Supervisory Committee: **Raja Mohd Kamil Raja Ahmad**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACOFSS	–	Ant Colony Optimization Feature Subset Selection
AICC	–	Akaike Information Criterion Corrected
ANN	–	Artificial Neural Network
CCD	–	Centroid Counter Distance
FFNN	–	Feed Forward Neural Network
GA	–	Genetic Algorithm
GAFNN	–	Genetic Algorithm and Feed Forward Neural Network
GLCM	–	Gray-Level Co-Occurrence Matrix
ICA	–	Independent Component Analysis
IDSC	–	Inner Distance Shape Context
KPCA	–	Kernel Principal Component Analysis
LVQ	–	Learning Vector Quantization
MAP	–	Maximum A Prior
MDL	–	Minimum Description Length
MMC	–	Move Median Center
MOO	–	Multi Objective Optimization
MOPGA	–	Multi Objective Parallel Genetic Algorithm
MSE	–	Mean Square Error
OLSA	–	Orthogonal Least Square Algorithm
PCA	–	Principle Component Analysis
PGA	–	Parallel Genetic Algorithm
PNN	–	Probabilistic Neural Network
PRS	–	Plant Recognition System
RBF	–	Radial Basis Function
RBNN	–	Radial Basis Neural Network
RFNN	–	Radial Function Neural Network
RLSA	–	Recursive Least Square Algorithms
ROI	–	Region of Interest
SIFT	–	Scale Invariant Features Transform
SMPD	–	Statistical Minimized Point Distance
SOM	–	Self-Organizing Map
SVM	–	Support Vector Machine

CHAPTER 1
INTRODUCTION

Plant Recognition System (PRS) is designed to assist botanists and, even, non-experts in identifying plants. Conventionally, botanists applied daunting techniques like vegetation key to identify right plant species. In this method, they needed to consider many characters such as shape, margin and venation in a plant leaf to correctly recognize a kind of plant. This method shows futile when dealing with the flora of a country or, even, a continent, which are mega-data [1]. Nowadays, intelligent systems are engaged to facilitate plant identification. Non-stop performance, portability and fastness of these systems caused replacement of them with the conventional methods [2]. Therefore, there is a demand to develop automated PRS to add more convenience and accuracy to it [3]. PRS is an intelligent system that benefits from machine learning and image processing techniques. Also, in some works, optimization methods have been used to improve the performance of PRS [4, 5]. The most prominent merit of this technique is convenient and fast implementation. Also, another benefit of this method is its automatic character. The main contribution of the optimization to PRS has been in the selection of the suitable features from feature vector to omit undesirable information [6, 7]. However, works (i.e., [8]), also, used this technique to optimize processes for training classifier. The focus of this thesis is on the improvement of the classification rate and automatic tuning of classifier by engagement of the optimization technique.

Several attempts have been made to improve the PSR. Broadly speaking, PRS involves data collection, image processing and feature extraction and classification. Each part of the PRS carries the same share of importance, meaning that selection of suitable or unsuitable methods in each part can increase or decrease the system performance. Therefore, the system performance can be improved by contribution in different parts. Conventional PRS didn't recruit optimization methods. SVM, neural network, linear discriminate analysis and back propagation neural networks are strong classifiers used in optimized PRS [2, 9, 10, 11]. In addition, classifiers like KNN, Probabilistic Neural Network (PNN), MMC and hyper sphere classification method has been applied to conventionally classification of plants [1, 5, 12]. However, some classifiers including statistical analysis, feature matching and combined classifier techniques showed weak performance [13, 14, 1, 15].

Also, the recruitment of the optimization method in PRS has been received increasing attention, in the recent research. Most of these works focus on the optimization of the feature vectors that are exorbitant. Methods such as Ant Colony Optimization (ACO) and Kernel Principle Component Analysis (KPC), which were used with SVM, that are applied to feature subset selection showed prominent classification rate [7, 6]. However, the approaches such as OLSA, RLSA and GA were used to optimize classifiers training [8, 4]. Overall, methods that used optimized PRS showed better performance compared to the non-optimized PRS. Therefore, the optimization techniques should be a part of PRS for improvement of classification rate.

1.1 Problem Statement

In the lack of botanists, computer aided systems are the best choice in plant identification. It is frequently proved that even a botanist expert cannot perform his works without the assistance of computers, especially when they are dealing with the huge dataset [13]. This drawback brought an increasing need for automated plant identification systems. Userfriendly, quick performance and reliability are the prominent benefits of this system. Existing PRS suffer from the choice of the best model for the classifiers. Most of them applied rules of thumb or trial and error method to find the classifier structures and parameters. There is no evidence whether these techniques can really help with tuning the classifier. This measure, also, can assist improving the classification rate. To sum up, problems are listed below:

1. Limitations of the existing PRS, such as lack of automatic classifier tuning and accuracy issue, bring the need for improving the current system.
2. Handling and processing huge plant's data in a country or, even, a continent necessitate developing fast and accurate PRS.
3. Portability and flexibility of this system allow everyone to have a nonstop expert, which is not a case with botanist.

From this point of view, improving the classification rate has been regarded considerably.

1.2 Objectives

Overall, the aim of this project is the development of the optimized PRS, through investigation of optimization of the classifier structure. To obtain this achievement, specific objectives are listed follow:

1. To design an automated PRS
2. To utilize Multi Objective Parallel Genetic Algorithm (MOPGA) to find the best set of structures for the Feed Forward Neural Network (FFNN)
3. To improve the classification rate applying tuned FFNN

1.3 Scopes of the Work

In the past, botanists and non-experts used key vegetation to identify the plants. This technique was very time-consuming and daunting because they needed to find the similar plant images through a large handbook to identify a right specific plant. Due to increasing automated and intelligent systems and the convenience they serve, people get more interested in using PRS. The core of this system is artificial intelligence and pattern recognition. In the PRS based on the plants' leaves images, first, plants images were collected using image acquisition devices such as scanners and cameras. Therefore, adequate number of plants images including Flavia, Citrus and Coleus were collected. Then, data were characterized by image processing and feature extraction techniques to be prepared for the next part of classification. Shape and texture based were two features applied for data characterization. In addition, machine learning techniques were applied to model the processed data. Therefore, for the last step of PRS, FFNN, which was chosen for machine learning technique, was tuned by MOPGA and then applied for classification of extracted features.

1.4 Thesis Layout

The thesis structure represents the stages of the PRS development based on the optimization method. The layout of the thesis is as below:

Chapter 2 reviews the recent growths in the PRS that employed machine learning and focuses on the conventional and optimized methods. Different methods such as Fast

Fourier Transform (FFT), geometrical wavelet for the feature extraction and Move Median Center (MMC) and FFNN for the classification and Orthogonal Least Square Algorithm (OLSA) and Genetic Algorithm (GA) optimization.

Chapter 3 provides the information about the research methodology. Initially, image data collection and feature extractions are covered. The last section of this chapter gives some insights into the evolutionary optimization techniques to tune FFNN and classification of the obtained data.

Chapter 4 analyses the method and the achievement of the research. Data collection, feature extraction, tuning the FFNN and classification results are shown. Then the results of the research are presented and discussed.

Chapter 5 concludes from the research findings and highlights of the results explores of the open problems. In the last part includes some additional suggestions for improvement of the future research.

REFERENCES

[1] J.-X. Du, X.-F. Wang, and G.-J. Zhang, "Leaf shape based plant species recognition," *Applied Mathematics and Computation*, vol. 185, no. 2, pp. 883–893, Feb. 2007.

[2] L. Zhang, J. Kong, X. Zeng, and J. Ren, "Plant species identification based on neural network," in *Proc. Fourth International Conference on Natural Computation*, 2008, vol. 5, pp. 90–94.

[3] S. Wu, "Flavia, a leaf recognition algorithm for plant classification using Probabilistic Neural Network (PNN)," 2007. [Online]. Available: http://flavia.sourceforge.net/. [Accessed: 14-May-2012].

[4] N. Valliammal and S. N. Geethalakshmi, "An optimal feature subset selection for leaf analysis," *International Journal of Computer and Communication Engineering*, vol. 6, pp. 406–411, 2012.

[5] S. G. Wu, F. S. Bao, E. Y. Xu, Y. Wang, Y. Chang, and Q. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," in *Proc. IEEE International Symposium on Signal Processing and Information Technology*, 2007, pp. 11–16.

[6] J. Tian, Q. Hu, X. Ma, and M. Han, "An improved KPCA/GA-SVM classification model for plant leaf disease recognition," *Journal of Computational Information Systems*, vol. 8, no. 18, pp. 7737–7745, 2012.

[7] X. Li and Z. Chen, "Weed identification based on shape features and ant colony optimization algorithm," in *Proc. International Conference on Computer Application and System Modeling*, 2010, vol. 1, pp. 39–50.

[8] J. Du, D. Huang, X. Wang, and X. Gu, "Shape recognition based on radial basis probabilistic neural network and application to plant species identification," in *Proc. Advances in Neural Networks*, 2005, vol. 3497, no. 2, pp. 281–285.

[9] F. Ahmed, H. A. Al-Mamun, A. S. M. H. Bari, E. Hossain, and P. Kwan, "Classification of crops and weeds from digital images: A support vector machine approach," *Crop Protection.*, vol. 40, pp. 98–104, Oct. 2012.

[10] J. Liu, S. Zhang, and S. Deng, "A method of plant classification based on wavelet transforms and support vector machines," *Emerging Intelligent Computing Technology and Applications*, pp. 253–260, 2009.

[11] S. Prasad, P. Kumar, and R. C. Tripathi, "Plant leaf species identification using curvelet transform," in *Proc. 2nd International Conference on Computer and Communication Technology*, 2011, pp. 646–652.

[12] Z. Wang, B. Lu, Z. Chi, and D. Feng, "Leaf image classification with shape context and SIFT descriptors," in *Proc. International Conference on Digital Image Computing Techniques and Applications*, 2011, pp. 650–654.

[13] Y. Ye, C. Chen, C.T. Li, H. Fu, and Z. Chi, "A computerized plant species recognition system," in *Proc. International Symposium on Intelligent Multimedia, Video, and Speech Processing*, 2004, pp. 723–726.

[14] L. Gao, X. Lin, M. Zhong, and J. Zeng, "A neural network classifier based on prior evolution and iterative approximation used for leaf recognition," in *Proc. 2010 Sixth International Conference on Natural Computation (ICNC)*, 2010, vol. 2 , pp. 1038–1043.

[15] J. C. Neto, G. E. Meyer, D. D. Jones, and A. K. Samal, "Plant species identification using elliptic Fourier leaf shape analysis," *Computers and Electronics in Agriculture*, vol. 50, no. 2, pp. 121–134, Feb. 2006.

[16] Z. Husin, A. Y. M. Shakaff, A. H. A. Aziz, R. S. M. Farook, M. N. Jaafar, U. Hashim, and A. Harun, "Embedded portable device for herb leaves recognition using image processing techniques and neural network algorithm," *Computers and Electronics in Agricalture*, vol. 89, pp. 18–29, Nov. 2012.

[17] M. Z. Rashad, B. S. El-Desouky, and M. S. Khawasik, "Plants images classification based on textural features using combined classifier," *International Journal of Computer Science & Information Technology*, vol. 3, no. 4, pp. 93–100, 2011.

[18] M. Shabanzade, M. Zahedi, and S. A. Aghvami, "Combination of local descriptors and global features for leaf recognition," *Signal and Image Processing: An International Journal (SIPIJ).* vol. 2, no. 3, pp. 23–31, 2011.

[19] A. Camargo and J. S. Smith, "Image pattern classification for the identification of disease causing agents in plants," *Computers and Electronics in Agriculture*, vol. 66, no. 2, pp. 121–125, May. 2009.

[20] S. T. Bow, *Pattern recognition and image preprocessing*, 2th ed. CRC Press, 2002, p. 720.

[21] C. M. Bishop, *Pattern recognition and machine learning*, vol. 1. Springer New York, 2006, p. 738.

[22] J. P. Marques De Sa, *Pattern recognition: concepts, methods, and applications*, 4th ed. Springer New York, 2001, p. 318.

[23] M. Negnevitsky, *Artificial intelligence: A guide to intelligent systems*. Addison Wesley/Pearson, 2011, p. 479.

[24] A. Koschan and M. Abidi, *Digital color image processing*. John Wiley & Sons, 2008, p. 450.

[25] M. Nixon, M. S. Nixon, and A. S. Aguado, *Feature extraction & image processing for computer vision*. Academic Press, 2012, p. 609.

[26] B. Yegnanarayana, *Artificial neural networks*. PHI Learning Pvt. Ltd, 2009, p. 476.

[27] J. Heaton, *Introduction to neural networks with Java*. Heaton Research, 2008, p. 440.

[28] T. Kumar and K. Verma, "A Theory based on conversion of RGB image to gray image," *International Journal of Computer Applications*, vol. 7, no. 2, pp. 7–10, 2010.

[29] L. Rutkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, *Artificial intelligence and soft computing, part I: The 10th International Conference on Artificial Intelligence and Soft Computing , Zakopane, Poland, June13-17,* 10th ed. Springer New York, 2010, p. 710.

[30] C. R. Tosh and G. D. Ruxton, *Modelling perception with artificial neural networks*. Cambridge University Press, 2010, p. 398.

[31] I. Mihajlović, Đ. Nikolić, N. Štrbac, and Ž. Živković, "Statistical modelling in ecological management using the artificial neural networks (ANNs)," *Serbian Journal of Management*, vol. 5, no. 1, pp. 39–50, 2010.

[32] M. Affenzeller, S. Wagner, S. Winkler, and A. Beham, *Genetic algorithms and genetic programming: modern concepts and practical applications*. Taylor & Francis, 2009, p. 379.

[33] S. N. Sivanandam and S. N. Deepa, *Introduction to genetic algorithms*. Springer, 2007, p. 462.

[34] B. Cyganek, *Object detection and recognition in digital images: theory and practice*. John Wiley & Sons, 2013, p. 552.

[35] G. Panchal, A. Ganatra, Y. P. Kosta, and D. Panchal, "Searching most efficient neural network architecture using Akaike's Information Criterion (AIC)," *International Journal of Computer Applications*, vol. 1, no. 5, pp. 41–44, 2010.

[36] P. C. Naval Jr and J. P. T. Yusiong, "An evolutionary multi-objective neural network optimizer with bias-based pruning heuristic," in *Proc. Advances in Neural Networks*, Springer, 2007, pp. 174–183.

[37] M. Gen, O. Katai, B. McKay, A. Namatame, R. A. Sarker, and B. T. Zhang, *Intelligent and evolutionary systems*, 4th ed. Springer, 2010, p. 218.

[38] R. C. Gonzalez, *Digital image processing using Matlab*, 12th ed. McGraw-Hill Education (India) Pvt Limited, 2010, p. 738.

[39] F. Albregtsen, "Statistical texture measures computed from gray level co-occurrence matrices," *Image Process. Lab. Dep. Informatics, University. Oslo*, pp. 1–14, Jan. 2008.

[40] B. Shrestha, "Classification of plants using images of their leaves," M.S. thesis, Dept. Computer. Sci., Appalachian State Univ., Carolina, U.S., 2010.

[41] C. Arunpriya and A. S. Thanamani, “An Efficient leaf recognition algorithm for plant classification using support vector machine,” in *Proc. International Conference on Pattern Recognition, Informatics and Medical Engineering*, 2012, pp. 428–432.

[42] C. Pandolfi, S. Mugnai, E. Azzarello, S. Bergamasco, E. Masi, and S. Mancuso, “Artificial neural networks as a tool for plant identification: a case study on Vietnamese tea accessions,” *Euphytica*, vol. 166, no. 3, pp. 411–421, 2009.

[43] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*, 5th ed. John Wiley & Sons, 2010, p. 784.

[44] E. G. M. De Lacerda, “Model selection of RBF networks via genetic algorithms,” Ph. D, Informatics Center of Pernambuco Federal University, Brazil March, 2003.

[45] J. B. Kadane and N. A. Lazar, “Methods and criteria for model selection,” *Journal of the American Statistical Association*., vol. 99, no. 465, pp. 279–290, 2004.