CLASSIFICATION MODELING FOR MALAYSIAN BLOOMING FLOWER IMAGES USING NEURAL NETWORKS

MUHAMMAD ASHRAQ SALAHUDDIN

MASTER OF SCIENCE (INFORMATION TECHNOLOGY)
UNIVERSITI UTARA MALAYSIA
2013
Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner in whole or in part, for scholarly purposes may be granted by my supervisor(s) or in their absence by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Kedah DarulAman
Malaysia
Abstrak

Pemprosesan imej merupakan bidang penyelidikan dalam bidang sains komputer yang berkembang pesat sejak ia diperkenalkan dan sehingga kini ia masih merupakan satu masalah yang mencabar dalam bidang multimedia dan perkomputeran. Bagi imej bunga-bungaan, masalah utama adalah disebabkan persamaan yang ketara di antara setiap bunga dengan bunga lain dari segi warna dan tekstur. Rupa bentuk imej seperti pencahayaan yang berbeza, kesan bayangan terhadap permukaan objek, saiz, bentuk, putaran dan kedudukan, latar belakang imej, keadaan bunga yang mengembang atau mula mengembang merupakan masalah yang dihadapi dalam pemprosesan imej. Oleh kerana pengecaman nama bunga masih lagi kompleks, kajian ini dijalankan bertujuan untuk membangunkan model pengelasan bunga bungaan Malaysia yang mengembang dengan menggunakan Rangkaian Neural berasaskan backpropagation. Imej bunga diekstrak melalui Region of Interest (ROI) di mana nilai warna dan tekstur sahaja yang akan diberi penekanan dalam kajian ini. Sejumlah 960 imej telah diekstrak daripada 16 jenis bunga. Setiap jenis bunga diwakili oleh 60 sampel ROI, manakala setiap ROI diwakili oleh 3 sifat warna (Hue, Saturation dan Value) dan 4 sifat tekstur (Contrast, Correlation, Energy dan Homogeneity). Menerusi fasa latihan dan pengujian, analisis dilaksanakan untuk meninjau prestasi rangkaian neural berdasarkan difficult to learn pattern yang digandakan (dirujuk sebagai DOUBLE) kerana ia mungkin dapat memberi gambaran mengapa imej bunga sukar untuk dikelaskan. Dapatan kajian menunjukkan bahawa Rangkaian Neural berdasarkan DOUBLE memperoleh ketepatan sebanyak 96.3% dan data asal sebanyak 68.3% manakala ketepatan Regresi Logistik dengan data asal ialah 60.5%. Hasil pengkelasan pohon pemutusan menunjukkan jChi-Squared Automatic Interaction Detection (CHAID) dan Extended Chi-Squared Automatic Interaction Detection (EX-CHAID) memperoleh prestasi yang paling tinggi sehingga mencapai 42% dengan DOUBLE. Dapatan kajian menunjukkan bahawa Rangkaian Neural dengan set data DOUBLE memperoleh prestasi tertinggi berbanding Regresi Logistik dan Pohon Pemutusan. Oleh itu Rangkaian Neural mempunyai potensi dalam pembangunan model bunga-bungaan Malaysia. Kajian pada masa akan datang boleh menumpukan kepada penambahan saiz data kajian dan ROI yang mungkin dapat meningkatkan prestasi ketepatan. Model klasifikasi bunga yang dibangunkan dalam kajian ini boleh dijadikan sebahagian daripada sistem pengecaman bunga Malaysia pada masa akan datang di mana warna dan tekstur diperlukan dalam proses pengecaman bunga.

Kata kunci: Klasifikasi imej, Rangkaian neural, Perseptron berbilang lapisan, Pohon pemutusan, Regresi logistik.
Abstract

Image processing is a rapidly growing research area of computer science and remains as a challenging problem within the computer vision fields. For the classification of flower images, the problem is mainly due to the huge similarities in terms of colour and texture. The appearance of the image itself such as variation of lights due to different lighting condition, shadow effect on the object’s surface, size, shape, rotation and position, background clutter, states of blooming or budding may affect the utilized classification techniques. This study aims to develop a classification model for Malaysian blooming flowers using neural network with the back propagation algorithms. The flower image is extracted through Region of Interest (ROI) in which texture and colour are emphasized in this study. In this research, a total of 960 images were extracted from 16 types of flowers. Each ROI was represented by three colour attributes (Hue, Saturation, and Value) and four textures attribute (Contrast, Correlation, Energy and Homogeneity). In training and testing phases, experiments were carried out to observe the classification performance of Neural Networks with duplication of difficult pattern to learn (referred to as DOUBLE) as this could possibly explain as to why some flower images were difficult to learn by classifiers. Results show that the overall performance of Neural Network with DOUBLE is 96.3% while actual data set is 68.3%, and the accuracy obtained from Logistic Regression with actual data set is 60.5%. The Decision Tree classification results indicate that the highest performance obtained by Chi-Squared Automatic Interaction Detection(CHAID) and Exhaustive CHAID (EX-CHAID) is merely 42% with DOUBLE. The findings from this study indicate that Neural Network with DOUBLE data set produces highest performance compared to Logistic Regression and Decision Tree. Therefore, NN has been potential in building Malaysian blooming flower model. Future studies can be focused on increasing the sample size and ROI thus may lead to a higher percentage of accuracy. Nevertheless, the developed flower model can be used as part of the Malaysian Blooming Flower recognition system in the future where the colours and texture are needed in the flower identification process.

Keywords: Image classification, Neural networks, Multilayer perceptron, Decision tree, Logistic regression.
Acknowledgement

Allamduillah. My greatest gratitude to Allah SWT for giving me the chance, time and ability to complete this research.

Foremost, I would like to express my sincere gratitude to my supervisor, Associate Professor Fadzilah Siraj for her invaluable guidance, encouragement and knowledge-sharing in completing this research. I am very thankful for her help, time, contributions and efforts in providing me all the guidance and constructive suggestion during this research. Not to forget, Associate Professor Abdul Nasir Zulkifli as my second supervisor.

I am also very grateful to my parents, Salahuddin Ahmad Tajuddin and Nasihah Mat Isa also to my wife Nur Syarima Ghazali for their sacrifices, helps, supports and prayers they have given. Without their encouragements, I will not able to continue study at this level.
# Table of Contents

Permission to Use ................................................................................................................................................. i
Abstrak.................................................................................................................................................................... iii
Abstract..................................................................................................................................................................... iv
Acknowledgement ...................................................................................................................................................... iv
Table of Contents ...................................................................................................................................................... vi
List of Tables .......................................................................................................................................................... viii
List of Figures ........................................................................................................................................................ x

## CHAPTER ONE INTRODUCTION ......................................................... 1
1.1 Overview ....................................................................................................................................................... 1
1.2 Problem Statements ....................................................................................................................................... 5
1.3 Objectives ...................................................................................................................................................... 8
1.4 Research Question ....................................................................................................................................... 9
1.5 Scope of the research .................................................................................................................................. 9
1.6 Significance of the research ........................................................................................................................ 9
1.7 Thesis Overview ........................................................................................................................................ 10

## CHAPTER TWO LITERATURE REVIEW .......................................... 11
2.1 Image processing ......................................................................................................................................... 11
2.2 Feature Extraction in Image Processing ..................................................................................................... 13
  2.2.1 Colour ....................................................................................................................................................... 15
  2.2.2 Texture .................................................................................................................................................... 18
2.3 Classification ................................................................................................................................................. 20
2.4 Artificial Intelligence Classifiers for Image Classification ........................................................................... 23
  2.4.1 Artificial Neural Networks ...................................................................................................................... 23
  2.4.2 Applications of ANN .............................................................................................................................. 26
  2.4.3 Multilayer Perceptron (MLP) ................................................................................................................. 28
  2.4.4 MLP in Pattern Recognition ................................................................................................................ 26
2.5 Logistic Regression ....................................................................................................................................... 32
2.6 Decision Tree ................................................................................................................................................. 35
  2.6.1 Decision Tree in Image Processing and Image Classification ............................................................. 36
2.7 Previous researches on flower image classification .................................................................................... 37
2.8 Summary ...................................................................................................................................................... 40
CHAPTER THREE METHODOLOGY ..................................................41
3.1 Introduction .................................................................................41
3.2 Methodology ............................................................................41
   3.2.1 Phase 1: Business Understanding ..............................................45
   3.2.2 Phase 2: Data Understanding ....................................................45
   3.2.3 Phase 3: Data Preparation ........................................................46
   3.2.4 Phase 4: Modeling .................................................................56
   3.2.5 Phase 5: Evaluation .................................................................59

CHAPTER FOUR RESULTS ....................................................................60
4.1 Preliminary Study Using NN .........................................................60
4.2 Experimental Results .................................................................73
4.3 Logistic Regression .....................................................................79
4.4 Decision Tree ............................................................................87
4.5 Summary ....................................................................................92

CHAPTER FIVE CONCLUSION AND RECOMMENDATION .......................93
5.1 Conclusion ...............................................................................93
5.2 Recommendations ....................................................................94
   5.2.1 Increase the number of Malaysian flowers datasets ...............94
   5.2.2 Include shape feature .............................................................94
   5.2.3 Environment .......................................................................105

APPENDIX .....................................................................................120
List of Tables

Table 2.1: Categories of image features................................................................. 14
Table 2.2: The selected features in previous flower image classification.................... 39
Table 3.1: Research Framework .............................................................................. 43
Table 3.2: RGB to HSV conversion formula.............................................................. 53
Table 3.3: Example of HSV colour space value ....................................................... 54
Table 3.4: Equation for Contrast, Correlation, Energy and Homogeneity of GLCM...... 55
Table 3.5: Examples of Contrast, Correlation Energy and Homogeneity datasets ......... 55
Table 3.6: Colour and Texture extraction values ....................................................... 56
Table 4.1: Flower types and images ......................................................................... 61
Table 4.2: Results of Hidden layer with 1800 flower images..................................... 63
Table 4.3: Scale Conjugate Gradient Training Result Architecture (Data Allocation: 80:10:10; Accuracy: 9.5%) ................................................................. 64
Table 4.4: Gradient Descent Training Result Architecture (Data Allocation: 80:10:10; Accuracy: 4.1%) ........................................................................ 65
Table 4.5: Original image and Misclassified image (Training; Data Allocation: 80:10:10, Accuracy: 9.5%) ................................................................. 66
Table 4.6: Original image and Misclassified image (Testing; Data Allocation: 80:10:10, Accuracy: 10.2%) ................................................................. 67
Table 4.7: Original image and Misclassified image (Training; Data Allocation: 80:10:10, Accuracy: 4.1%) ................................................................. 68
Table 4.8: Original image and Misclassified image ................................................... 69
Table 4.9: Reduction of flower dataset .................................................................... 71
Table 4.10: The 7 groups of flowers ....................................................................... 72
Table 4.11: Experimental results various hidden units and data allocation ............... 73
Table 4.12: Results of Hidden layer for Double repetition of hard pattern to learn ...... 75
Table 4.13: Results of Hidden layer for Triple repetition of hard pattern to learn ......... 76
Table 4.14: 60:20:20 scale conjugate ...................................................................... 77
Table 4.15: 60:20:20 Gradient Descents ................................................................. 78
Table 4.16: Logistic Regression Case Processing Summary ....................................... 79
Table 4.17: Logistic Regression Classification Result .............................................. 80
Table 4.18: Model Fitting Information ................................................................. 83
Table 4.19: Goodness of Fit................................................................................... 83
Table 4.20: Pseudo R-Square

Table 4.21: Likelihood Ratio Tests

Table 4.22: Parameter Estimates

Table 4.23: Decision Tree with original dataset of hard pattern to learn

Table 4.24: Cross Validation and Split Sample Validation

Table 4.25: Decision Tree with double repetition dataset of hard pattern to learn
List of Figures

Figure 1.1: Allamanda’s images taken under various lighting condition ..................................5
Figure 1.2: English Daisy’s flower images taken from different viewpoints ..........................6
Figure 1.3: Background clutter for different flowers is similar to each other ..........................6
Figure 1.4: Same flower but different state of blooming ....................................................7
Figure 2.1: RGB colour model ..............................................................................................15
Figure 2.2: CMYK colour model ............................................................................................16
Figure 2.3: YUV colour model ..............................................................................................16
Figure 2.4: Example of Colour Histogram .............................................................................17
Figure 2.5: Artificial Neural Networks ...................................................................................24
Figure 2.6: Architecture of Multilayer Perceptrons .................................................................27
Figure 2.7: Decision Tree Diagram .........................................................................................35
Figure 3.1: Theoretical framework of the research .................................................................42
Figure 3.2: Theoretical Framework .........................................................................................44
Figure 3.3: Samples of flower image taken ............................................................................46
Figure 3.4: RGB colour space to HSV colour space to grayscale colour space ..................49
Figure 3.5: Image Thresholding and Image morphology .........................................................50
Figure 3.6: Flower image without background .......................................................................51
Figure 3.7: MATLAB code for feature extraction .................................................................52
Figure 3.8: Texture extraction using MATLAB .....................................................................56
Figure 3.9: Steps in carrying out the experiment ....................................................................58
Figure 3.10: Original dataset has been split into Easy To Learn and Hard To Learn pattern .................................................................58
Figure 4.1: Flower dataset example by Nilsback (2009) .........................................................62
Figure 4.2: Data Allocation 60:20:20 (Scale Conjugate) .........................................................77
Figure 4.3: Data Allocation 60:20:20 (Gradient Descents) .....................................................78
Figure 4.4a: Flower Accuracy Percentage using Logistic Regression (Original) ..............81
Figure 4.4b: Flower Accuracy Percentage using Logistic Regression (Double repetition) ..82
Figure 4.4c: Flower Accuracy Percentage using Logistic Regression (Triple repetition) ....82
Figure 4.5: Decision Tree with Original dataset of hard pattern to learn ...............................87
Figure 4.6: Decision Tree with double repetition dataset of hard pattern to learn .................88
Figure 4.7: Decision Tree with Triple repetition dataset of hard pattern to learn ...................89
Figure 4.8: Decision Tree EX CHAID original flower dataset .............................................90

x
Figure 4.9: Decision Tree EX CHAID double repetition flower dataset ........................................ 91
Figure 4.10: Decision Tree EX CHAID Triple repetition flower dataset ....................................... 91
List of Appendices

Figure 1: MLP Original Dataset 60:20:20 Scale Conjugate Gradient ........................................... 120
Figure 2: MLP Original Dataset 70:20:10 Scale Conjugate Gradient ............................................ 120
Figure 3: MLP Original Dataset 80:10:10 Scale Conjugate Gradient ............................................. 121
Figure 4: MLP Original Dataset 60:20:20 Gradient Descents ....................................................... 121
Figure 5: MLP Original Dataset 70:20:10 Gradient Descents ....................................................... 122
Figure 6: MLP Original Dataset 80:10:10 Gradient Descents ....................................................... 122
Figure 7: MLP Double Dataset 60:20:20 Scale Conjugate Gradient .............................................. 123
Figure 8: MLP Double Dataset 70:20:10 Scale Conjugate Gradient .............................................. 123
Figure 9: MLP Double Dataset 80:10:10 Scale Conjugate Gradient .............................................. 124
Figure 10: MLP Double Dataset 60:20:20 Gradient Descents ....................................................... 124
Figure 11: MLP Double Dataset 70:20:10 Gradient Descents ....................................................... 125
Figure 12: MLP Double Dataset 80:10:10 Gradient Descents ....................................................... 125
Figure 13: MLP Triple Dataset 60:20:20 Scale Conjugate Gradient .............................................. 126
Figure 14: MLP Triple Dataset 70:20:10 Scale Conjugate Gradient .............................................. 126
Figure 15: MLP Triple Dataset 80:10:10 Scale Conjugate Gradient .............................................. 127
Figure 16: MLP Triple Dataset 60:20:20 Gradient Descents ....................................................... 127
Figure 17: MLP Triple Dataset 70:20:10 Gradient Descents ....................................................... 128
Figure 18: MLP Triple Dataset 80:10:10 Gradient Descents ....................................................... 128
CHAPTER ONE
INTRODUCTION

This chapter presents the background of the project that focuses on Malaysian Blooming Flowers classification. The problem statements and objectives of the study are also mentioned in this chapter. In addition, the research questions are formulated, the research scope as well as the significant of the research also provided in this chapter.

1.1 Overview

Classification is an active research area in data mining which most frequently involve the decision making (Zhang, 2000). Classification aims to predict categorical class labels for new samples (Dehkordi & Shenassa, 2006). It involves the process of grouping objects (information) accordingly into their belonging classes or groups based on their characteristic (Qi & Davidson, 2009).

In classification, there are two main schemes that are commonly used namely Supervised and Unsupervised classification. Supervised classification is the process of using samples of known identity or training data to classify pixels of unknown identity (Riviera & Manian, 2008). The training data are used to train the classifier which is tested with testing samples to evaluate the accuracy of the classifier which in turn is testing using test samples to evaluate the accuracy of the classifier. Some of the most commonly used supervised classification methods are Maximum Likelihood, Minimum Distance, Mahalanobis Distance and Neural Networks.
The contents of the thesis is for internal user only
REFERENCES


International Conference on Biomedical Engineering and Informatics 2010 (BMEI2010), 1, 424-427.


