Neural network approaches to caricature generation

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:


Metadata Record: https://dspace.lboro.ac.uk/2134/34440

Publisher: © Ka Hang Lai

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Please note that fines are charged on ALL overdue items.
Neural Network Approaches to Caricature Generation

by

Ka Hang Lai

A doctoral thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science
Loughborough University
January 2007

© by Ka Hang Lai 2007

Supervisors: Prof. Paul W H Chung and Dr. Eran Edirisinghe
Director of Research: Dr. Chris Hinde
ABSTRACT

Neural Network Approaches to Caricature Generation

Ka Hang Lai, November 2006

A caricature is defined as a humorous drawing of a human facial figure that makes some of its distinct features appear exaggerated. It is easily observed that the exaggerations made by different artists on facial components are often different and are non-linear. This uniqueness of the exaggerations signifies the drawing style of an artist, but has unfortunately been ignored in the design of existing computer based automatic caricature generation systems. Nevertheless learning the unique drawing style and modelling the non-linear exaggerations distinct to an artist provide the key but a real challenge to the computer based automatic generation of professional caricature.

This Thesis proposes a face modelling framework that includes two novel face models, which are capable of representing human faces in caricature applications. The first, the Simple Face Model contains 46 feature points and enables a simple but low resolution facial description. The second, the Enhanced Face Model, provides a more detailed representation of individual facial components with a total of 143 feature points. Both face models are tailored to meet the requirements of the proposed automatic caricature generation algorithms.

The Thesis proceeds to propose two novel example-based caricature generation systems where a neural network is used to capture the drawing style of an artist and a
morphic tool is subsequently used to automatically create new caricatures. The first approach considers the entire human face as a single object and performs caricaturing of all facial components simultaneously proving that neural networks are capable of accomplishing the drawing style capturing task. The second approach focuses on handling facial components independently and subsequently combining them together to produce a caricature. It is shown that this approach further improves the accuracy of drawing style capture. Subjective experimental results and detailed statistical analysis are provided to demonstrate the effectiveness of the proposed approaches.
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors, Prof. Paul W. H. Chung and Dr. Eran Edirisinghe, for their continuous support and guidance throughout this research work. They provided me an invaluable opportunity to pursue my dream. I also thank my director of research, Dr. Chris Hinde, for his time, effort and suggestions.

I am as ever, indebted to my parents, Mr. Joe Ming Lai and Mrs. Yin Fong Ho for their immeasurable love and unconditional support – especially my father’s vision and my mother’s patience have been of great significance throughout my life. I also wish to thank my brother, Mr. Ka Yeung Lai for his inspiration and encouragement during the last three years. I am also profoundly thankful to Miss Zhen Li, for her love, faith and understanding during my study. I dedicate this Thesis to them and to my late grandmother, Mrs. Sam Lai, who is in my heart at every moment of my life.

Special thanks must go to Murali for his technical support and advice. I am also grateful to caricaturists Mr. Abuhelga, Mr. Yang and Mr. Pang for their contributions during data collection of the project. Moreover, to the Department of Computer Science, Loughborough University for providing me the research studentship.

Last but not least, my sincere thanks to all my colleagues of the Digital Imaging Research Group, Dhammike, Murali, Nijad, Alvin, Moi, Iffat, Harish, Rupesh, Bardia and Mohammad, for making my life in Loughborough memorable and enjoyable.

Ka Hang Lai

17th November, 2006
# Table of Contents

Title page  
Abstract  
Acknowledgements  
Table of contents  
List of figures and tables  
Citations to published work  
Abbreviations

## PART 1: Introduction, Review and Background Technologies

### CHAPTER 1 OVERVIEW

1.1 Introduction  
1.2 Problem statement  
1.3 Project motivation  
1.4 Objectives  
1.5 Approach  
1.6 Contributions  
1.7 Thesis organisation

### CHAPTER 2 INTRODUCTION TO CARICATURE

2.1 Introduction  
2.2 What is a caricature  
2.3 Why caricature  
2.4 History of caricature  
2.5 How to draw caricature  
2.6 Conclusions

### CHAPTER 3 REVIEW OF 2D CARICATURE GENERATION SYSTEMS

3.1 Introduction  
3.2 Review of 2D caricature generation systems
CHAPTER 4 INTRODUCTION AND REVIEW OF BACKGROUND TECHNOLOGIES

4.1 Introduction 52
4.2 What is a human face 52
4.2.1 Introduction to human face 53
4.2.2 Facial components 53
4.3 Review of 2D geometric face models 54
4.3.1 Parametric models 54
4.3.2 Physically based models 55
4.3.3 Feature point based models 55
4.3.4 MPEG-4 standard 56
4.3.5 Choosing a geometric face model for 2D caricaturing 58
4.4 Introduction to artificial intelligence 60
4.4.1 An overview of AI 60
4.4.2 The definition of AI 60
4.4.3 Different AI technologies 60
4.4.4 Artificial neural networks 61
4.5 Introduction to image processing technologies 67
4.5.1 An overview of digital images 67
4.5.2 Basic image processing technologies 67
4.5.3 Applications of image processing technologies 69
4.6 Chapter summary and conclusions 74

PART 2: Contributions of Research

CHAPTER 5 PROPOSED FACE MODELLING FRAMEWORK AND DATASET PREPARATION 76
5.1 Introduction 76
5.2 A novel geometric face model with 46 feature points 77
5.3 An enhanced geometric face model with 143 feature points 82
5.4 Comparing proposed face models with MPEG-4 87
5.5 Dataset Preparation 88
5.5.1 Face selections 88
5.5.2 Inviting artists to draw caricatures 89
5.5.3 Normalisations 90
5.5.4 Manual marking of feature points 94
5.5.5 Mean face generation 95
5.5.6 Summary 97
5.6 Chapter summary and conclusions 98

CHAPTER 6 ENTIRE FACE-BASED CARICATURE GENERATION APPROACH 99
6.1 Introduction 99
6.2 An overview 100
6.3 Relationships among original image, caricature and mean face 102
6.4 Entire face-based caricature generation approach 104
6.4.1 Artificial neural networks 104
6.4.2 Preparation of training and validation sets 105
6.4.3 Neural network architectures 108
6.4.4 Neural network training and validations 111
6.4.5 Mesh warping 112
6.5 Experimental results and analysis 113
6.5.1 The validation experiments of artist-1 115
6.5.2 The validation experiments of artist-2 117
6.5.3 The validation experiments of artist-3 120
6.5.4 Further analysis of results 123
6.6 Subjective test 128
6.7 Further experiments 137
6.7.1 Different number of neurons in the hidden layer 137
6.7.2 Different neural network training algorithms and parameters 138
6.7.3 The experimental results of the second cross-validation instances 139
6.7.4 Comparisons with benchmark 142
6.8 Chapter summary and conclusions 143
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Thesis organisation</td>
<td>7</td>
</tr>
<tr>
<td>2-1</td>
<td>An example of a photo-caricature pair of Albert Einstein</td>
<td>9</td>
</tr>
<tr>
<td>2-2</td>
<td>Original image and corresponding caricatures drawn by three different artists</td>
<td>10</td>
</tr>
<tr>
<td>2-3</td>
<td>Comic slave of Greek statuette</td>
<td>12</td>
</tr>
<tr>
<td>2-4</td>
<td>Distorted faces drawn by Leonardo da Vinci</td>
<td>12</td>
</tr>
<tr>
<td>2-5</td>
<td>Italian singer with his wife drawn by Annibale Carracci</td>
<td>12</td>
</tr>
<tr>
<td>3-1</td>
<td>Caricature drawn by Brennan’s Caricature Generator</td>
<td>18</td>
</tr>
<tr>
<td>3-2</td>
<td>The relationships of mean face, original face and caricature</td>
<td>19</td>
</tr>
<tr>
<td>3-3</td>
<td>A total of 340 tessellations are defined over face space</td>
<td>20</td>
</tr>
<tr>
<td>3-4</td>
<td>Photo and caricature pair generated by Benson and Perrett’s system</td>
<td>21</td>
</tr>
<tr>
<td>3-5</td>
<td>Caricatures drawn by PICASSO system</td>
<td>22</td>
</tr>
<tr>
<td>3-6</td>
<td>An example of SCVF and DCVF</td>
<td>24</td>
</tr>
<tr>
<td>3-7</td>
<td>Caricature generated by Albert Pujol’s system</td>
<td>24</td>
</tr>
<tr>
<td>3-8</td>
<td>System framework that adopts eigenspaces in caricature generation</td>
<td>25</td>
</tr>
<tr>
<td>3-9</td>
<td>Caricature generated by Kaneko’s system</td>
<td>25</td>
</tr>
<tr>
<td>3-10</td>
<td>Caricature generated by Luo’s system</td>
<td>26</td>
</tr>
<tr>
<td>3-11</td>
<td>Caricatures generated by Chiang’s system</td>
<td>28</td>
</tr>
<tr>
<td>3-12</td>
<td>Caricatures generated by Mo’s system</td>
<td>29</td>
</tr>
<tr>
<td>3-13</td>
<td>Caricature generated by Gooch et al.’s system</td>
<td>30</td>
</tr>
<tr>
<td>3-14</td>
<td>System framework of Liang’s system that consists of training and runtime phases</td>
<td>32</td>
</tr>
<tr>
<td>3-15</td>
<td>Caricatures generated by Liang et al.’s system</td>
<td>32</td>
</tr>
<tr>
<td>3-16</td>
<td>Facial caricature drawing process of Iwashita’s system</td>
<td>33</td>
</tr>
<tr>
<td>3-17</td>
<td>Parameters of features</td>
<td>35</td>
</tr>
<tr>
<td>3-18</td>
<td>Membership functions of adverbs for modification</td>
<td>35</td>
</tr>
<tr>
<td>3-19</td>
<td>Facial caricature of Prof. Zadeh before and after modification by Iwashita’s system</td>
<td>35</td>
</tr>
<tr>
<td>3-20</td>
<td>System configuration of Nishino’s system</td>
<td>36</td>
</tr>
</tbody>
</table>
3-21 Caricatures generated by Nishino’s system 37
3-22 Exaggeration examples with all personality terms of Nishino’s system 37
3-23 A photo-caricature pair drawn by a professional artist 40
3-24 Original image, mean face and caricature - relationships diagram 49
4-1 Facial components of a human face 53
4-2 MPEG-4 facial definition parameters 57
4-3 A simplified biological neuron 61
4-4 A schematic representation of an artificial neuron 62
4-5 Reducing the size of an image by bilinear interpolation 68
4-6 Nearest neighbour, bilinear interpolation and bicubic interpolation 69
4-7 An example of image rotation 70
4-8 A simple example of digital image warping 71
4-9 An example of average mesh calculation 72
5-1 MPEG-4 facial definition parameters 77
5-2 Modified eyebrows FDPs (SFM) 78
5-3 Modified mouth FDPs (SFM) 79
5-4 Modified eyes FDPs (SFM) 79
5-5 Modified nose FDPs (SFM) 80
5-6 The proposed Simple Face Model (SFM) 81
5-7 Modified eyebrows FDPs (EFM) 82
5-8 Modified mouth FDPs (EFM) 83
5-9 Modified eyes FDPs (EFM) 83
5-10 Modified nose FDPs (EFM) 84
5-11 Modified ears FDPs (EFM) 84
5-12 Modified face FDPs (EFM) 85
5-13 The proposed Enhanced Face Model (EFM) 86
5-14 Comparison of MPEG-4 and proposed face models 87
5-15 The final normalised dataset 91
5-16 Original and corresponding caricature with marked FDPs 94
5-17 An example of average face 95
5-18 A snapshot of the software, ‘Morpher’ 96
5-19 Final generated mean face from ten original images 96
5-20 Dataset preparation steps
6-1 The proposed entire face-based automatic caricature generation algorithm
6-2 Calculation of $\Delta S$ and $\Delta S'$
6-3(i) The first validation experiment of artist-1
6-3(ii) The second validation experiment of artist-1
6-4(i) The first validation experiment of artist-2
6-4(ii) The second validation experiment of artist-2
6-5(i) The first validation experiment of artist-3
6-5(ii) The second validation experiment of artist-3
6-6 Validation results of the proposed system when trained on the caricatures of artist-1
6-7 Validation results of the proposed system when trained on the caricatures of artist-2
6-8 Validation results of the proposed system when trained on the caricatures of artist-3
6-9(i) Subjective test of the first validation case
6-9(ii) Subjective test of the second validation case
6-10 Results of the subjective test (Entire face-based approach)
6-11 Standardised normal distributions
6-12 Experimental results of using different numbers of hidden neurons
6-13(i) The second cross-validation results of artist-1
6-13(ii) The second cross-validation results of artist-2
6-13(iii) The second cross-validation results of artist-3
6-14 Caricatures produced by the Picasso benchmark with $\Delta S' = 1.5 \times \Delta S$
7-1 The proposed facial component-based automatic caricature generation algorithm
7-2 An example of a caricature before and after facial component positioning
7-3(i) The first validation experiment of artist-1
7-3(ii) The second validation experiment of artist-1
7-4(i) The first validation experiment of artist-2
7-4(ii) The second validation experiment of artist-2 157
7-5(i) The first validation experiment of artist-3 158
7-5(ii) The second validation experiment of artist-3 159
7-6(i) Subjective test on artist-1's computer generated caricatures 162
7-6(ii) Subjective test on artist-2's computer generated caricatures 162
7-6(iii) Subjective test on artist-2's computer generated caricatures 163
7-7 Results of the subjective test that compares caricatures generated using the two approaches 164
7-8(i) The second cross-validation results of artist-1 167
7-8(ii) The second cross-validation results of artist-2 168
7-8(iii) The second cross-validation results of artist-3 169
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Face shape of Iwashita's system</td>
<td>34</td>
</tr>
<tr>
<td>3-2</td>
<td>Feature elements of a face</td>
<td>34</td>
</tr>
<tr>
<td>3-3</td>
<td>Feature terms of eyes</td>
<td>34</td>
</tr>
<tr>
<td>3-4</td>
<td>Parallel transformation distance of each adverb</td>
<td>34</td>
</tr>
<tr>
<td>3-5</td>
<td>The parameters of Nishino's face caricature drawing system</td>
<td>37</td>
</tr>
<tr>
<td>3-6</td>
<td>Comparison of the characteristics of existing caricature systems</td>
<td>50</td>
</tr>
<tr>
<td>5-1</td>
<td>Comparison of the number of FDPs of MPEG-4 and two proposed models</td>
<td>87</td>
</tr>
<tr>
<td>6-1</td>
<td>An example of a part of a training set</td>
<td>106</td>
</tr>
<tr>
<td>6-2</td>
<td>Neural network architecture and training parameters</td>
<td>110</td>
</tr>
<tr>
<td>6-3</td>
<td>Critical values of z for one and two-tailed tests</td>
<td>133</td>
</tr>
<tr>
<td>6-4</td>
<td>z score calculations of the participants who were able to make a match</td>
<td>135</td>
</tr>
<tr>
<td>7-1</td>
<td>Comparison of two face models</td>
<td>149</td>
</tr>
<tr>
<td>7-2</td>
<td>A comparison of neural network architectures used by the two caricature approaches</td>
<td>150</td>
</tr>
<tr>
<td>7-3</td>
<td>z score calculations of the participants who chose the caricatures generated by the facial component-based approach</td>
<td>166</td>
</tr>
</tbody>
</table>
CITATIONS TO PUBLISHED WORK

The work presented in Chapters 6 and 7, and some sections of Chapter 5, has been submitted for journal publication, and is under review as follows:


The content of Chapters 5 and 6, has been published in the following papers:


The content of Chapters 5 and 7, has been published in the following paper:


The work presented in Chapter 3, is to be submitted as:

**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>Two Dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>Three Dimensional</td>
</tr>
<tr>
<td>AD</td>
<td>Anno Domini</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>AMA</td>
<td>Abstract Muscle Actions</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Networks</td>
</tr>
<tr>
<td>AR</td>
<td>Aleix Martinez and Robert Benavente</td>
</tr>
<tr>
<td>ASM</td>
<td>Active Shape Modelling algorithm</td>
</tr>
<tr>
<td>AU</td>
<td>Action Unit</td>
</tr>
<tr>
<td>CE</td>
<td>Coefficient of Efficiency</td>
</tr>
<tr>
<td>DCVF</td>
<td>Dense Caricaturing transformation Vector Field</td>
</tr>
<tr>
<td>DSCA</td>
<td>Drawing Style Capture Algorithm</td>
</tr>
<tr>
<td>EDFM</td>
<td>Exaggerating Difference From Mean</td>
</tr>
<tr>
<td>EFM</td>
<td>Enhanced Face Model</td>
</tr>
<tr>
<td>FACS</td>
<td>Facial Action Coding System</td>
</tr>
<tr>
<td>FDPs</td>
<td>Facial Definition Parameters</td>
</tr>
<tr>
<td>FFG</td>
<td>Facial Feature Grid</td>
</tr>
<tr>
<td>IEC</td>
<td>International Electrotechnical Commission</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organisation for Standardisation</td>
</tr>
<tr>
<td>LMS</td>
<td>Least Mean Square</td>
</tr>
<tr>
<td>MPEG</td>
<td>Motion Picture Expert Group</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial Least Square</td>
</tr>
<tr>
<td>SCVF</td>
<td>Sparse Caricaturing transformation Vector Field</td>
</tr>
<tr>
<td>SE</td>
<td>Standard Error of the Estimate</td>
</tr>
<tr>
<td>SFM</td>
<td>Simple Face Model</td>
</tr>
<tr>
<td>SVMs</td>
<td>Support Vector Machines</td>
</tr>
</tbody>
</table>
Part One

Introduction, Review and Background Technologies
Chapter 1

Overview

1.1 Introduction

Caricature is an art that conveys humor or sarcasm to people via drawing human faces. The basic concept is to capture the essence of the unique features of a subject and exaggerate them graphically. These artistic alterations change the relative ratio of the subject’s facial features and give a deeper impression to viewers. As caricatures are common in our daily lives, their computer based generation has attracted considerable research interests.

This chapter gives an introduction to the Thesis. It is organised as follows:

1.2 Problem statement
1.3 Project motivation
1.4 Objectives
1.5 Approach
1.6 Contributions
1.7 Thesis organisation
Chapter 1. Overview

1.2 Problem Statement

The problem addressed in this Thesis is formulated as:

How is it possible to capture the drawing style of an artist and automatically generate caricatures by a computer, depicting the captured drawing style?

In the past three decades, significant efforts have been made in generating caricatures by computers resulting in a number of approaches been reported [2-19]. However, none of these state-of-the-art computer based systems has attempted to learn the drawing style of an artist by using artificial intelligence technologies. As a result, a considerable amount of work remains to be done in this area.

1.3 Project Motivation

Though the ability to recognise a person in a caricature widely exists amongst humans, the ability of drawing caricature only exists in a few people, and has thus been considered as an innate talent. From the artists’ point of view, it is difficult for them to explain the craft of caricature generation as the associated drawing rules are embedded in their minds and often appear to be fuzzy [1]. Hence, developing an automatic caricature system with the ability to learn the mystery drawing rules of an artist based on his/her products is a significant challenge; especially as the underlying processes involved cannot be accurately explained. This also helps in understanding and comparing drawing styles of different artists in both artistic and psychological ways. Further, identification of caricatures drawn by a particular artist may become possible as his/her drawing rules are extracted and studied. Besides, such a system will find potential applications in situations where caricaturists require the help of a computer based system to generate artwork unique to their drawing styles with the intention of reducing workload.
1.4 Objectives
In order to tackle the problem identified above, this research project has two specific objectives:

1. To develop a novel face modelling framework that allows automatic caricaturing systems to generate caricatures based on numerical information.
2. To develop novel algorithms which are able to capture the drawing styles of different artists based on their caricature products and to automatically generate caricatures that are embedded with their styles.

1.5 Approach
The project was carried out in three phases:

1. Review the fundamental concepts of drawing caricature and carry out a survey of existing automatic caricaturing technologies.
2. Define a novel face modelling framework, which provides robust platforms for facial caricaturing.
3. Apply and adopt an artificial intelligence technology that is able to capture the drawing style from the caricature products of an artist.

An in-depth understanding of the process of drawing caricature and a complete awareness of the existing automatic caricature generation technologies are used as a basis for the whole project. The proposed framework and solutions have been built to address the major shortcomings of the current systems, and then generate caricatures.

Prior to the start of caricaturing, face modelling platforms suitable for representing
static facial figures and corresponding caricatures were established. Two novel geometric face models were developed to fulfill the needs of the proposed drawing style capture algorithms.

The application of an artificial intelligence technology, which is responsible for learning the drawing style of an artist and subsequently generates caricatures based on the captured rules, is the core of the proposed research. This process was supported by the face modelling framework provided above.

1.6 Contributions

The contributions of this Thesis are:

1. A comprehensive literature review on existing computer based automatic caricature generation techniques.

2. A face modelling framework that provides two platforms for numerical representation of human faces in caricaturing applications.

3. A novel caricaturing approach using neural networks that learns from entire face-based caricatures, which is able to learn and capture the drawing style of an artist based on his/her caricature products.

4. A novel approach of using neural networks in generating facial component-based caricatures, which is able to improve on the performance of the entire face-based algorithm while computational resource requirements are maintained.
1.7 Thesis Organisation

This Thesis is organised into two parts where each part focuses on a particular subject and can be read separately based on the reader's background and interest. The Thesis organisation is diagrammatically illustrated in figure 1-1.

Part 1: Introduction, Review and Background Technologies

Chapter 1 provides an overview of the project. It identifies the problem addressed in the Thesis, specifies the motivation and objectives. It further discusses the adopted approach and highlights the contributions. Lastly, it outlines the Thesis organisation.

Chapter 2 gives an introduction to the art of caricaturing. It describes what a caricature is, why people need caricatures, the history and how to draw caricatures.

Chapter 3 reviews different computer based caricaturing systems. The advantages and disadvantages of each approach are critically compared and discussed.

Chapter 4 gives an introduction to the background technologies used within the project. It starts with a review of different geometric face models. It further introduces an artificial intelligence technology - neural network, and discusses its capabilities of capturing the drawing style of an artist. It proceeds to discuss the image processing techniques used within the proposed caricaturing system, such as morphing and warping.

Part 2: Contributions of Research

Chapter 5 proposes a novel face modelling framework, with two different geometric face models, which provides platforms to the following proposed drawing style...
Chapter 1. Overview

capture algorithms. It further describes the steps of dataset preparation.

Chapter 6 proposes an entire face-based caricature generation approach to capture the
drawing style of an artist with the aid of a neural network. It proceeds to provide
experimental results, detailed analysis and validations.

Chapter 7 proposes a facial component-based caricature generation approach to
capture the drawing style of an artist by using a neural network, and includes
experimental results, detailed analysis and validations.

Finally, Chapter 8 concludes the Thesis with a summary of contributions, practical
challenges, potential applications and future directions of research.
Chapter 1. Overview

Readers with interest in current caricature systems, face models, basic AI technologies and image processing techniques

Chapter 2
Introduction to caricature

Chapter 3
Review of 2D caricature generation systems

Chapter 4
Introduction and review of background technologies

Chapter 5
Proposed face modelling framework and dataset preparation

Chapter 6
Entire face-based caricature generation approach

Chapter 7
Facial component-based caricature generation approach

Chapter 8
Thesis summary, conclusions and future work

Figure 1-1. Thesis organisation.
Chapter 2

Introduction to Caricature

2.1 Introduction

Caricaturing has become a part of art in human culture with great popularity at present. It is common to see artists drawing caricatures at popular tourist destinations. Political caricatures appear in newspapers while caricatures of celebrities are frequently found in magazines.

This chapter gives an introduction to caricature. It is organised as follows:

2.2 What is a caricature?
2.3 Why caricature?
2.4 History of caricature
2.5 How to draw caricature
2.6 Conclusions
2.2 What is a Caricature?

A caricature is a portrait of a subject with deliberate exaggerations to subject's essence in order to produce a humorous effect, which also conveys a message to the readers in an artistic way [1]. However, not everybody is capable of drawing caricature as the inborn talent only exists in a few people.

The word "Caricature" originated from the Italian word Caricare, which means to load or exaggerate [20]. According to the Oxford English Dictionary [21], caricature is defined as:

"A funny drawing of someone that makes some of their features look bigger or more amusing than they really are."

Caricature has always been confused with the word "Cartoon". Although both of them are freehand drawings, the main difference is that cartoons have no facial exaggerations of a subject but caricatures have. Besides, a cartoon can be a drawing of any object but a caricature is restricted to a portrait of a human face only.

The following is a typical example of a photo-caricature pair [22]:

Figure 2-1. An example of a photo (left) – caricature (right) pair of Albert Einstein.
2.3 Why Caricature?

People enjoy caricatures and are also interested in recognising facial figures from corresponding caricatures. Even if a caricature is drawn by simple lines, it has enough characteristic information to be recognised. Thus, caricatures convey messages in a simple and efficient way [3]. The main reason for the popularity of caricature is that the underlying message expressed by the artist via drawing is usually humorous, ridicule or satirical.

However, the caricatures of the same person created by different artists can be very different, since the artists' drawing styles play an important role [23]. Although the same subject from both caricatures is still able to recognised, the detailed messages conveyed from them may not be the same. Figure 2-2 illustrates an example of the same image caricatured by different artists, where the caricatures show diverse results.

![Figure 2-2. Original image (a) and corresponding caricatures (b), (c) and (d) drawn by three different artists. Note that for the same subject (a), the nose and face lengths of the caricatures from different artists can be very diverse.](image)

The caricaturing process is based on deforming the features of a face selectively. It creates images by exaggerating the most distinctive facial features and distorting the common parts or leaving them unchanged. Hence, the information in the caricature is richer and more compact than the original image, which means that a caricature is
more effective in face recognition. This has been verified by psychological experiments that have proved the time required for recognising caricatured images is less than the corresponding veridical ones [24-26].

Face distinctiveness effects are significant in human face processing theories [27-29]. Studies have demonstrated that faces rated as distinct by observers are easier to be recognised than faces rated as typical [30]. As a result, psychologists are interested in caricature generation systems as they provide a platform to explore the distinctiveness effects in both face perception and recognition research [24,26,31]. Further application of caricatures in face recognition and psychology studies will be discussed in Chapter 8.


2.4 History of Caricature

Caricature has a very long history. The earliest example goes back to 79 A.D., where a Roman soldier scratched some centurions on the wall of a camp in Pompeii. Evidence also suggests that the Greeks (Figure 2-3) and Romans enjoyed embedding ridicule messages in humorous drawings [38].

In the fifteenth century, the work of Leonardo da Vinci can equally well be referred to as caricaturing. A careful investigation of distorted faces drawn by him reveals that they are an artist's experiments in drawing portraits with different forms of ugliness and expressions, rather than portraits of true subjects [38]. (Figure 2-4)

However the inventor of modern caricature, Annibale Carracci, was only recognised in 1560. Annibale was not only credited for the invention of the art but also of the word "Caricature". He was reported to have said, "A good caricature, like every work of art, is more true to life than reality itself" [38]. (Figure 2-5)
2.5 How to Draw Caricature

Caricaturists have an admirable ability to draw humorous human faces by capturing their distinguishable facial features and exaggerating them. Although drawing caricatures depend on the innate magical talents of the artists, artists such as Hughes and Redman [1,23] published books to provide general guidelines for beginners.

Besides, Akleman [39] developed a procedure to make caricatures using an interactive morphing tool, which is suitable for users without any innate caricaturing talent. The software adopted a very simple algorithm that utilised a 2D deformation tool to generate caricature according to the following steps:

1. Begin with an extremely simple template where the essential features are represented by a few number of lines.
2. Exaggerate only one feature at a time.
3. Exaggerate the feature in one direction, if it does not create a likeness, try exaggerating in the opposite direction.
4. Repeat the same procedure on another feature until the result is satisfactory.

No matter if an artist is a beginner or professional, the road to success in drawing caricature, is the same. In 1991, Benson and Perrett [3] listed the main keys to caricature generation:

1. Who the person is and our familiarity with him/her.
2. The distinctiveness of particular features of that face.
3. Which features are exaggerated and the degree to which this is done.
4. The artist’s style and bias.
5. The sociopolitical climate (current affairs and public sentiment).
2.6 Conclusions

This chapter provided an introduction to caricaturing. It first, described what a caricature is and provided a definition. It then proceeded to explain how caricatures help in scientific research. It further introduced the history of caricature. After that, the procedures adopted for drawing caricature by beginners and artists without innate caricaturing talent were discussed. Finally, the chapter listed the key criteria of drawing successful caricatures.

Caricature has a very long history in human culture, which is defined as an art of pictorial drawing with the subject's distinctive features deliberately exaggerated, in order to give a better impression than in the original image. These make caricature suitable for psychological researchers in understanding the perceptive ability of human brains. Caricature also provides a channel for the artists to convey amusing messages to readers, which are usually simple and easy to understand. Although not everyone has the talent to draw caricatures, following the guidelines in books or using a 2D interactive tool, it is possible for non-artists to create caricatures of reasonable quality.
Chapter 3

Review of 2D Caricature Generation Systems

3.1 Introduction

Computer based automatic caricature generation has gained increased popularity as a research topic in the last thirty years [2-19]. It is a combination of several areas of research that includes computer science, psychology and human biology. Researchers have been making efforts to invent different caricature generation systems. With the aid of various computer technologies, automatic generation of high quality caricature has become increasingly possible.

This chapter reviews and discusses the caricature generation systems proposed in previous literature. It is organised as follows:

3.2 Review of 2D caricature generation systems

3.3 Discussion

3.4 Conclusions
3.2 Review of 2D Caricature Generation Systems

In the last three decades, twelve major caricature generation systems have been reported in literature [2-4,9-17]. Though each system has its inherent advantages and disadvantages, none of them has been able to generate professional caricature of photorealistic nature. The reasons for their limited performance come from different aspects, which will be fully discussed in this chapter.

The existing caricature generation systems can be broadly classified into geometric and linguistic approaches. The geometric approach is a more popular choice with ten systems being reported [2-4,9-15], whereas only two systems have attempted to generate caricature using the linguistic approach [16-17]. However regardless of the above mentioned presentation method used by the caricature generation systems, most share the same underlying caricaturing principle.

3.2.1 Geometric Caricature Generation Systems

The geometric approach generally performs more efficiently than the linguistic approach, due to its ability of capturing the subtle features of facial components accurately.

Brennan’s “Caricature Generator” [2]

Susan E. Brennan [2], 1975, is the first person who attempted to develop a computer-assisted 2D caricature generation system. The success of her system aroused interest in researchers within multidisciplinary research backgrounds. This well-known breakthrough became a benchmark for the algorithms that followed and has been widely used in both psychological and face recognition experiments [24,31-33]. Brennan developed a simple algorithm in her system which consists of
three important steps: marking of facial feature points, generating a mean face and then exaggerating the original image.

(i) Marking of facial feature points

Brennan is a forerunner of research who adopted the feature point-based face representation. 160 feature points have been chosen as an effective set of parameters that can be used to represent a human face [2]. All the facial feature points are marked manually in Brennan’s “Caricature Generator”. A number of systems introduced thereafter were based on the design of Brennan’s facial feature points [3-4,13].

(ii) Generating a mean face

The human brain has a remarkable inborn ability of remembering and recognising thousands of faces it encounters during a lifetime, despite the fact that most of the faces are metrically similar. To this effect, Psychologists have suggested that everybody has a mental picture of a “mean face” within their mind [31-32], which is an average of the faces encountered during lifetime. It is widely accepted that the memorisation of a given face is then carried out by storing only the distinctive features as compared to the mean face [34-37]. This explains the phenomenon where faces with more distinctive features being easily recognisable.

Brennan adopted and implemented the mean face concept into her “Caricature generator”. The mean face was generated using ten Caucasian males as most of the target subjects were Caucasian males. Brennan proposed that the subject and the mean face should share certain general characteristics, which can help in generating more accurate results.
(iii) Exaggerating the original image

Caricature is considered a rendering that emphasises the distinctive features of a particular face. A formalisation of this idea, Exaggerating the Difference From the Mean (EDFM), is widely accepted among psychologists and caricaturists to be the driving factor behind caricature generation [23,31].

"Caricature generator" accepts the feature points of the original face and the mean face as the inputs, a normalisation process is then provoked which scales the original face according to the distance between the pupils of the mean face, as Brennan assumes the distance between pupils is the same for any faces. The system will then determine the difference between each point on the face and its corresponding point on the mean face. A caricature can be generated by increasing the distance between the feature point coordinate in the original from those of the mean face. Finally, the new position of each point will be calculated and curves will be defined by cubic B-splines [40]. An example of a caricature generated by Brennan's system is shown in figure 3-1.

![Figure 3-1. Caricature drawn by Brennan's Caricature Generator [2].](image)

(iv) The relationships of mean face, original face and caricature

Psychologists introduced a "face space" framework for representing human faces in 1991 [29]. Faces are thought of as points in a multidimensional space, as the distance...
between any two points is served as a measure of the similarity between the faces. The relationships of mean face, original face and caricature are represented in figure 3-2.

![Figure 3-2. The relationships of mean face, original face and caricature [41].](image)

Caricature enhances the distinctiveness of a particular face by moving the point of the original face away from the mean face in a face space, which is made along the line connecting the original face to the mean face.

**Benson and Perrett’s System [3]**

Benson and Perrett [3] extended Brennan’s system in 1991; they modified Brennan’s approach by allowing the feature translation to be applied on the input image directly, in order to produce a photorealistic caricature. The results are more impressive than Brennan’s system as photographs always provide more information than line drawings. Therefore, the generated caricatures are easier to be recognised [3].

Benson and Perrett increased the number of facial feature points to 186 based on Brennan’s design. The additional points are mainly around the eyes and the mouth as Ellis and Young [42-43] reported that internal features are most significant for identification of a familiar face. Besides, Benson and Perrett suggested that successful
caricaturing depends on distributing the markers evenly along a feature edge or contour while still capturing the distinctive shape of the feature.

On top of this, Benson and Perrett applied the face mesh technology [44] to their caricature generation system. The basic idea of constructing a face mesh is tessellating the image surface into a set of triangles based on the facial feature points, as shown in figure 3-3.

![Figure 3-3. A total of 340 tessellations are defined over face space [3].](image)

In Benson and Perrett’s system, 340 triangular tessellations were adopted to map an original image. A full mapping algorithm can be found in Benson and Perrett’s work, “Synthesising continuous-tone caricatures” [3]. Once the mapping has been done, the image can be exaggerated or distorted by simply moving the facial feature points into the desired shape. Within this process the location of each pixel in the image is recalculated automatically. The users can alter the degree of exaggeration interactively. While a positive distortion forms a caricature, a negative distortion reduces all deviations between the veridical and mean faces and forms an anti-caricature, which is an image that moves towards the mean face in a face space. Although anti-caricatures are rarely used in the art of caricaturing, it has applications in face recognition (see section 8.4). An example of caricature generated by Benson and
Chapter 3. Review of 2D Caricature Generation Systems

Perrett’s system is shown in figure 3-4.

![Original image and Caricature](image)

Figure 3-4. Photo and caricature pair generated by Benson and Perrett’s system [3].

**Koshimizu’s PICASSO System [4]**

PICASSO is another well-known caricature generation system, it was introduced by Koshimizu et al. for generating 2D facial caricatures in 1999 [4]. They adopted the same mean face approach proposed by Brennan [2]. However, the PICASSO system is reported as the first fully automatic caricature generation system. Koshimizu developed an image processing module in his system to extract different facial components. Firstly, irises were extracted using Hough transform algorithm [45], and then facial regions of eyebrows, nose and mouth were extracted according to a profile. Finally, the contour of each facial component was extracted by K-L Expansion [46]. The boundaries of all facial components are represented by a set of facial feature points, which are then compared with the mean face to generate caricature. Some caricatures generated by PICASSO system are shown in figure 3-5.
Koshimizu et al. further proposed lots of interesting caricature generation systems based on the work of PICASSO. For instance, the same idea was implemented into a 3D basis [5] using a range cameras to capture the 3D face data, and then use the same mean face principle described above to generate 3D caricatures.

Moreover, Koshimizu et al. introduced the popular Motion PICASSO system [6]. In this system, due to the fact that a still caricature is not sufficient to express individual facial character, motion features of emotional expressions are added to enforce facial caricaturing. With the aid of an image recorder, sequential caricatures are encoded into frames of a video.

Besides the above systems, Koshimizu et al. further developed the Interactive PICASSO [7] system. They suggested that a caricature generation system should consider how human vision extracts the feature points and recognises them. In the traditional approach, caricaturing only happens between the image and the caricaturist, with the flow of information always treated as one way, from the caricaturist to the image. However, caricatures of a subject vary from caricaturist to caricaturist.
Similarly, the evaluation varies from viewer to viewer, i.e., though one may say that a particular caricature is 'good', another may disagree. Therefore, an interactive mechanism is required between the viewer and the image. Koshimizu et al. equipped the PICASSO system with an eye-camera, which can mount on the user’s head to capture the viewpoint distributions when an image is displayed to the viewer. The system calculates the exaggeration rate of each facial component according to the viewpoint distributions. The most viewed component receives the highest exaggeration rate as the more viewed part is believed to be more distinct than the other components. As a result, the caricature generated under this feedback mechanism is expected to be the best caricature for the viewer concerned.

Recently, Koshimizu et al. introduced the Web PICASSO system [8]. It is a PICASSO system implemented on a website, which is open to the worldwide Internet users. Any visitor can join the facial caricaturing process in real time. Since the principle of PICASSO is significantly based on “mean face”, all visitors are initially invited to upload their facial images, which will contribute towards the computation of the ultimate mean face of the world.

Albert Pujol et al.’s System [9]

A new approach was proposed by Albert Pujol et al. in 2000 [9]. Though the traditional mean face concept was adopted within their system, the step of manual feature point marking procedure was replaced by using two transformation vector fields. The first vector field adopted is the Sparse Caricaturing Transformation Vector Field (SCVF) [47], which is responsible for capturing the geometric differences between the original image and the mean face. The SCVF construction algorithm encodes how the valley points of the face image should be deformed to obtain a
caricaturised version with respect to the mean face. Then the SCVF is applied to the whole face image to develop a Dense Caricaturing Transformation Vector Field (DCVF) [47], which provides a smoother and more detailed deformation than the SCVF. Finally, the DCVF is applied to the original face image to generate a caricature. By using these two fields, the whole caricaturing process can be done automatically without marking any facial feature points. An example of SCVF and DCVF are shown in figure 3-6. The resulting caricature generated by Albert Pujol’s system is shown in figure 3-7.

![Figure 3-6. An example of SCVF (left) and DCVF (right) [9]. The shadings represent valley points respect to the mean face.](image)

![Figure 3-7. Caricature (right) generated by Albert Pujol’s system [9].](image)

**Kaneko’s System [10]**

Kaneko introduced a facial caricature generator using Eigenspaces in 2002 [10]. Once again, the mean face concept introduced by Brennan [2] was implemented. However, the representation of faces is different from the traditional facial feature point approach. Kaneko used the technology of principal component analysis (PCA) [48] to generate eigenvalues and eigenvector for each facial component, such as eyebrows, eyes, nose, mouth and frontal face contour. The difference between a component and the mean face can be simply calculated by subtracting the eigenvectors of them, i.e. the comparisons of different facial components become simple.
In the proposed system, Kaneko emphasised that the shapes of facial components and their arrangement should not be treated together. It is because the variety of representation may be restricted and causes an unfavorable collapse in the synthesised image when the degree of exaggeration becomes large. As a result, eigenspaces are derived for the shape of each facial part and the arrangement of facial parts separately. Exaggeration process is performed on each eigenspace independently in order to obtain a high flexibility while generating caricatures. Finally, by combining the results obtained from the exaggerated shape of each facial part and the exaggerated arrangement of facial parts, a caricature can be generated. The framework of the caricature process and the example of caricatures generated by Kaneko's system are shown in figure 3-8 and 3-9 respectively.

Figure 3-8. System framework that adopts eigenspaces in caricature generation [10].

Figure 3-9. Caricature generated by Kaneko's system [10]. (a) Original image. (b) Initial contour shape. (c) Mean face. (d) Only shapes are emphasised. (e) Both shapes and arrangement are emphasised. (f) Final facial caricature with hair.
Chapter 3. Review of 2D Caricature Generation Systems

Luo's System [11]

In 2002, Luo proposed a caricature generation system with embedded exaggeration rules [11], which are also based on Brennan’s mean face concept. For the reason of simplicity, all facial feature points on the input images are marked manually. Then the contour of each facial component is calculated using a Berzier curve [40]. Each facial component has its own rules and is treated individually. The following is an example of an exaggeration rule of a face:

“If someone’s forehead is taller / shorter, we increase / decrease the distance between the hairline and the brow.”

The comparisons of different facial components with a mean face or standard are not described in the paper and the implementation details are unknown. Luo simply referred to the exaggeration process adopted, as, “if we feel that the nose is big, we can exaggerate it by using the bottom of the nose as the central point and increasing the size proportionally”. All the exaggerated facial component layers will then be merged together to generating a complete caricature. A special feature of this system is that Luo added the shading and coloring layers to the final output based on his own rules, which increased the reality of caricatures generated by the system. An example of caricaturing using Luo’s approach is illustrated in figure 3-10.

Figure 3-10. Caricature generated by Luo’s system [11]. (a) Input picture. (b) Output picture without exaggeration. (c) Output picture with exaggerated nose and mouth.
Chapter 3. Review of 2D Caricature Generation Systems

Chiang’s System [12]

With the emergence of the MPEG-4 standard in the late 1990s, Chiang introduced a caricature generation system based on the Facial Definition Parameter (FDP) set defined with the standard [49-51]. Chiang adopted the 119 facial feature points derived from the FDPs instead of the 160 facial feature points suggested by Brennan [2] in 1975. Once again, Chiang applied Brennan’s mean face concept. 100 photographs of Asian females were collected and used for the computation of a mean face. Subsequently a statistical analysis was adopted to find out the normal range, average and standard deviation of each facial component. Chiang adopted a simple linear model for assigning the exaggeration rate outside the normal range. This simple algorithm is defined as follows:

\[
\text{If (value within normal range)} \\
\quad \text{value remains unchanged} \\
\text{Else if (value < range_min)} \\
\quad \text{value} = \text{value} - (\text{range_min} - \text{value}) \times \text{scale} \\
\text{Else} \\
\quad \text{value} = \text{value} + (\text{value} - \text{range_max}) \times \text{scale}
\]

As the facial components exaggerate within a face mesh [52] based on the above algorithm, a photorealistic caricature will be generated. In addition, Chiang further applied the system to generate non-photorealistic caricatures. The subject (figure 3-11a) is first drawn by an artist to create a facial portrait (figure 3-11c), to which the modified feature point positions of the generated photorealistic caricature (figure 3-11b) is applied, to obtain a non-photorealistic caricature (figure 3-11d). Examples of photorealistic and non-photorealistic caricatures are shown in figure 3-11.
Figure 3-11. Caricatures generated by Chiang’s system [12]. (a) A subject with the original face mesh overlaid. (b) Modified node positions after shape exaggeration. (c) An artist’s work. (d) Caricature generated based on (c). (e) Another artist’s work. (f) Caricature generated based on (e).

Mo’s System [13]

The mean face concept has been adopted by different caricaturing systems, and the exaggeration of the difference from the mean is widely accepted among caricaturists [1,23]. However, Mo suggested that exaggerating the difference from the mean might not always produce the best caricatures. The distinctiveness of a displaced feature does not only depend on its distance from the mean, but also the variance. Mo provided an example to support his argument [13]:

“The width of the mouth is much more widely distributed than the width of eyes. Thus, a mouth 2cm wider than the mean may still look normal, whereas eyes 2cm wider than the mean will be very distinctive. In the exaggeration of the difference from the mean method, both the mouth and eye width will be emphasised by a same factor because their difference-from-mean values are the same.”
In view of the above factor, Mo used a different way to represent faces. 300 facial images were collected and a matrix that stores both x and y coordinates of every facial feature shape was constructed. Then a non-negative matrix factorisation [53] procedure was adopted to compute the face space dimension. Each dimension consists of a basis vector and its distribution. When a new image is input to the system, its facial shape will be represented in the face space as a non-negative linear combination of the basis vectors and a residual. The shape is then exaggerated by a factor if the difference-from-mean value is larger than two standard deviations. The residual may consist of distinctive features that cannot be represented in the face space, thus, the residual is also exaggerated but with a reduced scale. Finally, photorealistic and non-photorealistic caricatures are generated by warping the image from the original shape to the exaggerated shape. Figure 3-12 shows the original image, photorealistic caricature and non-photorealistic caricatures generated by Mo’s system.

Figure 3-12. Caricatures generated by Mo’s system [13]. Left: Original faces. Middle: Photorealistic caricatures. Right: Stylised caricatures.
Gooch et al.'s System [14]

Gooch et al. presented a semi-automatic caricaturing system in 2004 [14]. They adopted the popular mean face approach in their system. First of all, they converted a target photograph into a black-and-white illustration [54]. Then they framed the illustration by four borderlines. After that, they introduced further four vertical lines to mark the inner and outer corners of the eyes respectively. Next, three additional horizontal lines were added to mark the position of the eyes, the tip of the nose, and the mouth. This set of horizontal and vertical lines is named as a Facial Feature Grid (FFG). Likewise, they generated a FFG for the mean face.

When a user frames an original image, the difference between the mean face grid and the user-set grid can be calculated by subtracting the corresponding vertices in both grids. Then the result will be scaled by a given percentage and the source image will be warped correspondingly. Besides, the exaggeration can be adjusted by modifying the FFG interactively. The FFG and the resulting caricature are shown in figure 3-13.

![Figure 3-13. Caricature generated by Gooch et al.'s system [14]. First: The face is framed by four borderlines. Second: Facial features and interior lines are matched. Third: Both grid and the underlying image are warped interactively. Fourth: The resulting caricature.](image)
Liang et al.’s System [15]

Liang et al. [15] proposed the first example-based caricature generation system in 2003, which is reported as a system that has the ability to learn from caricatures drawn by an artist, capture the artist’s understanding of what distinctive facial features are and also his/her exaggeration style. Liang et al. adopted a prototype-based learning and exaggeration model, in which the calculation of mean face is not necessary. The system handles the shape and texture separately, so it consists of two modules: shape exaggeration and texture style transferring.

(i) Shape exaggeration module

First of all, 92 original facial images and their corresponding caricatures drawn by an artist were prepared. For the simplicity of the learning process, all images considered included frontal views of facial images, without accessories. The system then applied the Partial Least Square algorithm (PLS) [55] to analyse these original image-caricature pairs, and subsequently summarised into 28 caricaturing prototypes. These prototypes represent different exaggeration patterns of the artist.

In the runtime phase, when a new image is input to the system, the contours of facial features will be extracted by an Active Shape Modelling algorithm (ASM) [56]. Then the extracted shape is classified into one of the defined prototypes, which matches the main facial features of the input image to the style which is most likely exaggerated by the artist. Finally, the contours of the facial features are exaggerated according to the selected prototype.

(ii) Texture style transferring module

Liang et al. adopted the example-based sketching system developed by Chen [57] for
Chapter 3. Review of 2D Caricature Generation Systems

Texture style transferring. It first converts the input image into a pencil sketch cartoon. Subsequently the cartoon is deformed into a caricature based on the modified feature point positions obtained from the shape exaggeration part.

The system framework and two examples of generated caricature are shown in figure 3-14 and 3-15 respectively.

Figure 3-14. System framework of Liang’s system that consists of training and runtime phases [15].

Figure 3-15. Caricatures generated by Liang et al.’s system [15].
3.2.2 Linguistic Caricature Generation Systems

The linguistic approach is a less popular choice since the accuracy of handling facial features is not as good as in the geometric approach. However, the linguistic approach provides a convenient way for users to generate caricatures using common human language.

Iwashita’s System [16]

In 1997, Iwashita introduced the first linguistic facial caricature generation system based on fuzzy theory [58], which is able to handle linguistic expression inputs. The system adopted the mean face concept suggested by Brennan and consisted of three important modules: input, output, and modification, as shown in figure 3-16.

![Figure 3-16. Facial caricature drawing process of Iwashita's system [16].](image)

(i) Input module

The input part deals with the face shape. The facial features include eyes, nose, mouth, eyebrows and ear are defined using linguistic expressions. Firstly, the user chooses the
sex and a face shape from table 3-1. Next, user has to choose a feature term with an adverb for each element of a feature as the input to the system. Table 3-2 lists the facial features and their elements. Table 3-3 lists the feature terms that corresponds to the facial feature (i.e. component), eyes. Finally table 3-4 lists all the adverbs and their corresponding transformation distances.

Table 3-1. Face shape of Iwashita’s system.

<table>
<thead>
<tr>
<th>Face shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A moon-shaped face</td>
</tr>
<tr>
<td>2. An egg-shaped face</td>
</tr>
<tr>
<td>3. A triangular-shaped face</td>
</tr>
<tr>
<td>4. A rectangular-shaped face</td>
</tr>
<tr>
<td>5. A home base shaped face</td>
</tr>
</tbody>
</table>

Table 3-2. Feature elements of a face.

<table>
<thead>
<tr>
<th>Features</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eyes</td>
<td>Size, slant, others</td>
</tr>
<tr>
<td>Nose</td>
<td>Size, height, size of wings of nose direction</td>
</tr>
<tr>
<td>Mouth</td>
<td>Size, thickness of lips</td>
</tr>
<tr>
<td>Eyebrows</td>
<td>Thickness, depth of color, length slant</td>
</tr>
<tr>
<td>Ears</td>
<td>Size</td>
</tr>
</tbody>
</table>

Table 3-3. Feature terms of eyes.

<table>
<thead>
<tr>
<th>Size</th>
<th>Slant</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big</td>
<td>Lower</td>
<td>Big pupils</td>
</tr>
<tr>
<td>Small</td>
<td>Upper</td>
<td>Small pupils</td>
</tr>
<tr>
<td>Thin</td>
<td></td>
<td>Laughing look in eyes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sleepy-looking in eyes</td>
</tr>
<tr>
<td>Normal*</td>
<td>Normal*</td>
<td>Normal*</td>
</tr>
</tbody>
</table>

Table 3-4. Parallel transformation distance of each adverb.

<table>
<thead>
<tr>
<th>Adverb</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Completely</td>
</tr>
<tr>
<td>2.</td>
<td>Very</td>
</tr>
<tr>
<td>3.</td>
<td>Fairly</td>
</tr>
<tr>
<td>4.</td>
<td>Slightly</td>
</tr>
<tr>
<td>5.</td>
<td>Only a little</td>
</tr>
</tbody>
</table>

* Normal means the impression of the feature in an average face.

(ii) Output module

Every facial feature has several parameters, as shown in figure 3-17, of which values are expressed by fuzzy sets and adjusted by adverbs. Each parameter has a value which is represented by a fuzzy set defined on the interval [-1,1] with a triangular typed membership function \( N(r-0.2, r, r+0.2)(-1 \leq r \leq 1) \), where the sign of \( r \) shows
the increase or the decrease of the parameter value from the mean face. The change of the parameter value is adjusted by an adverb according to its parallel transformation distance listed in table 3-4. Subsequently the parameter values are calculated based on their membership functions and are used to exaggerate the mean face into a caricature.

(iii) Modification module

The modification part accepts linguistic expressions to improve the unsatisfactory caricatures generated by the system. A new set of membership functions and adverbs as shown in figure 3-18 are used for modification. Figure 3-19 shows an example of generated caricatures before and after the modification.
Nishino's System [17]

Nishino proposed another linguistic caricature generation system in 1998 [17], which can be divided into a simple caricature drawing module and an exaggeration module with exaggeration rules embedded within. The system framework is illustrated in figure 3-20.

First of all, Nishino designed a face space with 21 parameters, as illustrated in table 3-5. Any face can be defined as a subset and represented in the face space. Once the face space has been defined, the process of generating one's caricature can be considered as seeking a point that is most nearest to his/her face configuration. Several caricaturists were invited to describe some provided caricatures with their own words. The results were used to compile the feature term table and define the fuzzy sets and membership functions. Hereafter, when a user chooses all the feature terms with an adverb either "very" or "slightly" for each of them, these linguistic expressions will be passed through the processes of fuzzification, intersection, aggregation and defuzzification [58] to generate a caricature which is represented by the face space.
Nishino further improved the system by adding an exaggeration module. He investigated the caricature products of the invited caricaturists and constructed 117 exaggeration rules, and then embedded these rules into the system. An exaggeration process is determined by a combination of feature terms and personality terms. For example, “delightful-personality” changes the feature term into a magnified form. The “big eyes” and the “small eyes” become the “huge (very big) eyes” and the “tiny (very small) eyes” [17]. A simple generated caricature and exaggeration examples with all personality terms are illustrated in figure 3-21 and 3-22 respectively.

<table>
<thead>
<tr>
<th>No.</th>
<th>param. name</th>
<th>range</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>p₁</td>
<td>pupil Size</td>
<td>0 ≤ p₁ ≤ 50</td>
<td>pixel</td>
</tr>
<tr>
<td>p₂</td>
<td>eye slant</td>
<td>-45 ≤ p₂ ≤ 45</td>
<td>degree</td>
</tr>
<tr>
<td>p₃</td>
<td>distance between both eyes</td>
<td>50 ≤ p₃ ≤ 150</td>
<td>pixel</td>
</tr>
<tr>
<td>p₄</td>
<td>eye size</td>
<td>40 ≤ p₄ ≤ 100</td>
<td>pixel</td>
</tr>
<tr>
<td>p₅</td>
<td>nose size</td>
<td>25 ≤ p₅ ≤ 150</td>
<td>pixel</td>
</tr>
<tr>
<td>p₆</td>
<td>nose height</td>
<td>15 ≤ p₆ ≤ 70</td>
<td>pixel</td>
</tr>
<tr>
<td>p₇</td>
<td>nose coordinate</td>
<td>0 ≤ p₇ ≤ 70</td>
<td>pixel</td>
</tr>
<tr>
<td>p₈</td>
<td>mouth size</td>
<td>20 ≤ p₈ ≤ 150</td>
<td>pixel</td>
</tr>
<tr>
<td>p₉</td>
<td>mouth distance from nose</td>
<td>5 ≤ p₉ ≤ 50</td>
<td>pixel</td>
</tr>
<tr>
<td>p₁₀</td>
<td>face width</td>
<td>0 ≤ p₁₀ ≤ 100</td>
<td>pixel</td>
</tr>
<tr>
<td>p₁₁</td>
<td>face length</td>
<td>20 ≤ p₁₁ ≤ 200</td>
<td>pixel</td>
</tr>
<tr>
<td>p₁₂</td>
<td>face shape coordinate</td>
<td>0 ≤ p₁₂ ≤ 40</td>
<td>ratio</td>
</tr>
<tr>
<td>p₁₃</td>
<td>distance between eye and eyebrow</td>
<td>0 ≤ p₁₃ ≤ 35</td>
<td>pixel</td>
</tr>
<tr>
<td>p₁₄</td>
<td>eyebrow slant</td>
<td>0 ≤ p₁₄ ≤ 50</td>
<td>pixel</td>
</tr>
<tr>
<td>p₁₅</td>
<td>hair length</td>
<td>80 ≤ p₁₅ ≤ 300</td>
<td>pixel</td>
</tr>
<tr>
<td>p₁₆</td>
<td>front hair length</td>
<td>0 ≤ p₁₆ ≤ 70</td>
<td>pixel</td>
</tr>
<tr>
<td>p₁₇</td>
<td>forehead size</td>
<td>-20 ≤ p₁₇ ≤ 35</td>
<td>pixel</td>
</tr>
<tr>
<td>p₁₈</td>
<td>lip width</td>
<td>0 ≤ p₁₈ ≤ 30</td>
<td>pixel</td>
</tr>
<tr>
<td>p₁₉</td>
<td>eyebrow width</td>
<td>0 ≤ p₁₉ ≤ 50</td>
<td>pixel</td>
</tr>
<tr>
<td>p₂₀</td>
<td>ear size</td>
<td>0 ≤ p₂₀ ≤ 100</td>
<td>pixel</td>
</tr>
<tr>
<td>p₂₁</td>
<td>glasses size</td>
<td>-10 ≤ p₂₁ ≤ 50</td>
<td>pixel</td>
</tr>
</tbody>
</table>

Table 3-5. The parameters of Nishino’s face caricature drawing system.
3.2.3 Commercial Caricature Machines

It is common to see commercial caricaturing booths in big theme parks. Although the underlying technologies of these commercial systems are likely to be one of the published caricaturing algorithms explained above or their variants, the exact algorithm used and their detailed operation have not been quoted by the manufacturers. Two of the existing commercial products have been briefly studied as follows:

"Digital Easel" by Magical Enterprises Inc. [18]

"Digital Easel" [18] is a machine which is able to generate caricature interactively. It first takes a photo of the participating user and then provides him/her a number of caricature templates to choose from, which are sample caricatures already drawn by a professional artist. The machine will then simply caricature the photo of the user according to the style of the chosen template.

"Foto Morph" by American Alpha Inc. [19]

"Foto Morph" [19] is a morphing machine rather than a caricaturing machine. It provides funny effects by combining two different faces together and then generating an average face. One such option is called "Gene Machine" that combines one male and one female face together and predicts the face of their offspring. A common approach for combining faces is known as "Morphing" [63], which will be further discussed in Chapter 4.
3.3 Discussion

Most of the existing caricature generation systems have been introduced in section 3.2. A critical review of their operational aspects and their strengths/weaknesses are discussed in this section:

3.3.1 Evaluation: Geometric Caricature Generation Systems

Brennan’s “Caricature Generator” [2]

Brennan is the pioneer of research in the area of automatic caricature generation. Her caricature generator has been widely applied in different areas proving its success and contributions [24,31-33]. The system adopted a very simple algorithm that imitates the entire process of drawing caricature in an artist’s mind. The fundamental idea of the system is that caricatures can be drawn by exaggerating a human face according to its differences with the mean face. As a result, the system applied the simplest methodologies in the preparation work, such as, manual marking of facial feature points that makes the process more accurate, and simple linear exaggeration in generating caricatures.

However, Brennan’s “Caricature Generator” has some limitations. Although the mean face theory is widely agreed by psychologists and caricaturists [23,31], the exaggeration part of the system may not truly reflect the way that artists express the differences between the original image and the mean face in caricatures. Figure 3-23 is a good example to explain this.
Chapter 3. Review of 2D Caricature Generation Systems

The above photo-caricature pair was drawn by a professional artist [22]. The right ear of the above subject clearly illustrates that a caricatured ear is not only exaggerated in size, but also non-linearly changed in shape and orientation. However, in Brennan’s “Caricature Generator”, the exaggeration part only provides linear exaggeration of facial components by scaling with a single factor, which means only the size can be changed. As a result, inadequacy of handling non-linear exaggerations will lead to a decrease in realism of the generated caricature.

Moreover, the exaggeration rate for the whole face is hard-coded in the Caricature Generator. However, the exaggeration rates used by a caricaturist various in different facial components, which are also affected by the shape and orientation. Therefore, the exaggeration rates are recommended to be defined from investigating a professional artist’s caricatures.

Benson and Perrett’s System [3]

The fundamental theory of Benson and Perrett’s system is entirely based on Brennan’s design and therefore the limitations of Brennan’s system [2] still apply. The novelty of this system is the final output module which provides impressive photographic quality
caricatures. However, the generation of photographic caricature is relatively computational expensive, which also increases the complexity of the system due to the fact that handling color and shading are more complicated than handling black-and-white simple lines, in the traditional approach. However, the face mesh methodology implemented in the system provided a useful platform for handling images. It completely isolates the image processing part from the system, which helps in testing the exaggeration algorithm with just a few moves of facial feature points.

Koshimizu's PICASSO System [4]
The basic PICASSO system is entirely based on Brennan's concept of a "Caricature Generator". The contribution of PICASSO system focuses on the automatic feature extraction part instead of exaggeration part. As a result, the limitations of Brennan's system discussed above apply to the PICASSO system and the quality of the generated caricatures remains unchanged.

However some extensions to the PICASSO system, such as the Interactive PICASSO system [7], introduced improved approaches to generating caricatures by considering interactions from viewers.

Albert Pujol et al.'s System [9]
Albert Pujol et al.'s system emphasises the use of two transformation vector fields to replace manual marking of facial feature points, in order to speed up the process of outlining different facial components. However, manual marking of facial feature points is recognised as a simple and accurate method, which also decreases the complexity and computation time of the system.
Chapter 3. Review of 2D Caricature Generation Systems

Kaneko’s System [10]

Kaneko adopted eigenspaces as the platform to represent faces in the proposed system. Similar to Albert Pujol et al.’s system [9], the traditional manual marking of facial feature points was abandoned. Again, this method accelerates the system by skipping manual facial feature extractions and also decreases the workload. However, it raises complexity, computation and accuracy related problems.

Luo’s System [11]

Luo took a different approach to performing exaggerations by defining all the exaggeration rules himself and forcibly embedding the drawing style within the system via the rules. The proposed system provides a solution to tackling the problem of non-linear exaggerations in Brennan’s system [2] as these drawing rules can be defined to exaggerate non-linearly. Furthermore, the exaggeration rate of each facial component can be defined independently, which solves another problem that exists in Brennan’s system. In view of the above, the accuracy of the exaggeration rules is critical and directly affects the quality of the generated caricature. Although consulting a professional caricaturist before defining the rules is highly recommended, due to reasons discussed in section 1.3, getting an artist to express these rules verbally is a difficult, if not impossible task. Further, in Luo’s paper the methods used to define exaggeration rules have not been quoted.

The layer concept introduced by Luo is an advantage of the system. Each facial component is considered as a standalone layer where exaggerations are independent to each layer. In other words, a fully component based caricature system was developed by Luo.
Chapter 3. Review of 2D Caricature Generation Systems

Chiang’s System [12]

The system proposed by Chiang is the first attempt to use the FDPs of MPEG-4 [49-51] as the facial feature points. The utilisation of this new authoritative standard is always recommended as it allows different systems to work on the same platform, which makes them comparable and compatible with each other.

Chiang defined the exaggeration algorithm based on measuring each facial component in a statistical way, which is another example of a component based caricature generation system as all the facial components are exaggerated independently, with their specific exaggeration rates.

The final stage of Chiang’s system applies the warping technique [63] to generate non-photorealistic caricatures. After selecting an artist’s work as the source image, exaggerations will be performed on it. This method simply converts an artist’s portrait into a caricature in order to illustrate the drawing style of this particular artist in the resulting caricature. However, there is no learning, understanding or style capturing from artist’s work, and a portrait of each new face has to be drawn by the artist before each caricature generation.

Mo’s System [13]

Mo provided a clear example to explain the weaknesses of the difference-from-mean exaggeration method, and proposed a mathematical solution to tackle the problem. In fact, the issue does not exist if each facial component is treated independently with its own exaggeration rate. The system introduced by Chiang [12] proposed a very simple statistical analysis to avoid the problem mentioned above, where each facial component is measured separately. The normal range, average and standard deviation
of each component are recorded and used to calculate the exaggeration rate. As a result, the problem of the difference-from-mean exaggeration method was overcome.

Gooch et al.'s System [14]

Gooch et al. proposed a novel facial feature grid to frame a human face, which resulted in the whole caricaturing process being simple and fast. However, Gooch et al. used only eleven lines to replace the 160 facial feature points in Brennan’s system [2], which highly decreases the accuracy of handling exaggerations of facial components and subtle features. Therefore, the caricature generated by Gooch et al.'s approach is only impressive when the contour of the face is exaggerated.

Liang et al.'s System [15]

Liang et al.’s system is the only caricature generator attempted to learn from an artist’s products. Liang et al. took a prototype-based learning and exaggeration approach to generate caricatures. However, the number of prototypes defined by their system is too small. Using 28 prototypes to represent all kinds of human faces in the world is definitely inadequate. Hence, inappropriate exaggeration occurs if there is no suitable prototype that fits the input image. Moreover, this prototype algorithm is not a component-based approach. The input face could have more than one distinctive facial features, such a face may be exaggerated in an undesirable way as finding a prototype that fulfills the required combination of distinctive facial features (within a limited 28 choices) is difficult. As a result, accurate caricatures cannot be generated.

In the caricature preparation stage, Liang et al. requested the artist to select only several key facial features to exaggerate and to maintain a consistent exaggeration rate throughout the drawings. However, any restrictions to the artist’s drawing effort are
Chapter 3. Review of 2D Caricature Generation Systems

inappropriate as the style of the artist will be seriously affected. Thus the drawing style in the caricatures used in Liang et al.'s work cannot represent the actual drawing style of this artist.

Furthermore, Liang et al. invited the artist to draw 92 caricatures for the project. However, asking an artist to draw so many caricatures is unreasonable and time consuming, which also decreases the applicability of the work. Besides, no scientific evaluation was carried out on the final outcomes of the proposed system as the presented evaluations were only based on visual comparisons by the authors and hence could be highly subjective and biased. Further, Liang et al. adopted a statistical approach - Partial Least Square (PLS) [55], which is an algorithm developed from Principal Component Analysis (PCA) [48], to accomplish the learning of the artist's drawing style. The use of various artificial intelligence technologies, such as neural networks, could have been considered as they provide more effective machine learning ability [75].

3.3.2 Evaluation: Linguistic Caricature Generation Systems

Iwashita's System [16]

Iwashita developed a linguistic system that implemented Brennan's mean face idea [2] in a linguistic manner. However, the system does not have the ability to exaggerate automatically. The caricature generated by Iwashita's system strictly followed the linguistic expressions input from the user. In other words the user has to find out the distinctive features of a face and subsequently instruct the system via linguistic expression inputs, where to exaggerate the face. Consequently, the system is very limited to those users with the talent of locating the distinctive features of a face, which greatly decreases the usability of a caricature generation system.
Nishino’s System [17]

Nishino introduced a linguistic caricature system with embedded exaggeration rules. There are two parts in Nishino’s system: the simple caricature drawing part and the exaggeration part. In the caricature drawing part, it has the same problem as Iwashita’s system [16] that does not have the ability to exaggerate automatically. In the exaggeration part, Nishino embedded the drawing rules into the system for generating exaggerated caricatures based on different personality terms. However, according to the definition of caricature in Chapter 2, a caricature is a portrait in which the distinctive features of the original are exaggerated. That means the exaggeration process focuses on the distinctive features only, no matter what kind the personality of the subject is. Hence, Nishino’s system is a personality drawing system instead of a caricature generation system.

3.3.3 Summary

After reviewing different caricaturing systems, it can be concluded that nearly all of them share the same caricaturing principle – the mean face theory, which exaggerates the distinctive features of an original image away from the calculated mean face by a scaling factor. This approach is widely agreed by caricaturists [1,23] and adopted in most caricaturing systems [2-14,16-17]. However, the limitation of the current caricaturing systems that implemented this concept is that the exaggerations are only performed in a linear manner. As shown in figure 3-23, there is no doubt that professional caricaturists not only scale the distinctive features, but also modify their shapes and orientations. The inadequacy of handling non-linear exaggerations is the main reason that causes the products of the current caricature systems to be non-professional and unsatisfactory.
Furthermore, most of the systems applied the same scaling factor to all distinctive facial components. However, professional caricaturists exaggerate different facial components with various amounts. Besides, different caricaturists have different exaggeration rules in their mind, which define their unique drawing styles (see section 2.3). As a result, a caricaturing system should be capable of caricaturing facial components individually, which allows each component to be exaggerated using its own exaggeration rate.

In considering the above problems faced by the existing caricature generation systems, learning from the products of a caricaturist is a possible solution to find out the exaggeration rules embedded in an artist's mind that govern their unique drawing style. Subsequently, applying these extracted rules to a new image and converting it into a caricature is the most effective way to imitate the drawing style of a particular artist. As a result, the fundamental aim of the research presented in this Thesis is to capture the drawing style of a given artist from his/her caricaturing products and subsequently generating caricatures with embedded drawing style. However, the quality of the generated caricatures is totally depended on the caricaturing products provided for experimentation. If the artist is professional, the generated caricatures are expected to be professional and hence include a distinct drawing style. Yet, if the artist is amateur, the outputs will be amateur and hence less likely to have a unique style. Therefore, the caricatures generated by the proposed system can be considered as a medium to present the drawing style of the particular artist, which means the objective of this project is slightly different from that of the existing caricaturing systems that aimed to generate the best quality caricatures.

Due to the reason that the drawing products of the caricaturists are in 2D, it is a more
appropriate platform for the proposed work as compared to using a 3D caricaturing platform that used in some existing work [5,59-60]. Analysing the drawing style from an artist’s products can further be classified into shape modification and texture representation [15]. As the main concern of the project is to find out the exaggeration rules embedded in an artist’s brain, the texture analysis is ignored. The resulting caricatures adopt the photo-realistic approach instead of the freehand line-drawing approach. Converting a photo-realistic caricature into a line-drawing image can be done by using widely available commercial packages [61-62]. However, learning the texture representation of an artist from his/her caricaturing products can be a standalone research topic.

A closer analysis of the above caricature generation systems reveals that the geometric approach [2-15] is the main-stream in this research area, obviously due to the reason that it provides a precise facial feature extraction and modification platform. Although the linguistic approach [16-17] is more human understandable, it is difficult to describe the features of an object precisely, especially its shape. This also explains why caricaturists find it difficult to explain their inborn drawing rules to others. Therefore this project adopts the geometric approach rather than the linguistic approach.

Based on the above detailed analysis, the following diagram illustrates the relationships among an original image, the mean face and its corresponding caricature drawn by an artist [15]:

---
Chapter 3. Review of 2D Caricature Generation Systems

When an artist captures the differences ($\Delta S$) between the original image and the mean face in his/her mind, based on the original image, related changes ($\Delta S'$) are then made in the caricature. The target of this project is to capture the relationship between $\Delta S$ and $\Delta S'$ that defines the rules in the artist’s brain. More details will be discussed in Chapter 5.

The table 3-6 provides a comprehensive comparison of the characteristics of existing computer based caricaturing systems. It also compares the functionality of the proposed approach with that of the existing approaches.
### Table 3-6. Comparison of the characteristics of existing caricaturing systems.

<table>
<thead>
<tr>
<th>Caricature generation systems</th>
<th>Geometric or Linguistic Approach</th>
<th>No. of feature points</th>
<th>Automatic feature extraction</th>
<th>Adopt Mean Face (EDFM)</th>
<th>Output quality</th>
<th>Embed artist’s rules</th>
<th>Only exaggerate linearly</th>
<th>Same scaling factor to all components</th>
<th>Exaggeration module</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brennan</td>
<td>Geometric</td>
<td>160</td>
<td>No</td>
<td>Yes</td>
<td>Line-drawing</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Semi-automatic</td>
</tr>
<tr>
<td>Benson and Perrett</td>
<td>Geometric</td>
<td>186</td>
<td>No</td>
<td>Yes</td>
<td>Photo-realistic</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Semi-automatic</td>
</tr>
<tr>
<td>Koshimizu (Piscco)</td>
<td>Geometric</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Line-drawing</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Automatic</td>
</tr>
<tr>
<td>Albert et al.</td>
<td>Geometric</td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
<td>Photo-realistic</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Automatic</td>
</tr>
<tr>
<td>Kaneko</td>
<td>Geometric</td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
<td>Line-drawing</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Automatic</td>
</tr>
<tr>
<td>Luo</td>
<td>Geometric</td>
<td>Unknown</td>
<td>No</td>
<td>Yes</td>
<td>Line-drawing</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Automatic</td>
</tr>
<tr>
<td>Chiang</td>
<td>Geometric</td>
<td>119</td>
<td>No</td>
<td>Yes</td>
<td>Photo-realistic</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Automatic</td>
</tr>
<tr>
<td>Mo</td>
<td>Geometric</td>
<td>94</td>
<td>No</td>
<td>Yes</td>
<td>Both</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Automatic</td>
</tr>
<tr>
<td>Gooch et al.</td>
<td>Geometric</td>
<td>30</td>
<td>No</td>
<td>Yes</td>
<td>Photo-realistic</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Semi-automatic</td>
</tr>
<tr>
<td>Liang et al.</td>
<td>Geometric</td>
<td>-</td>
<td>Yes</td>
<td>No</td>
<td>Line-drawing</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Automatic</td>
</tr>
<tr>
<td>Iwashita</td>
<td>Linguistic</td>
<td>Unknown</td>
<td>No</td>
<td>Yes</td>
<td>Line-drawing</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Manual</td>
</tr>
<tr>
<td>Nishino</td>
<td>Linguistic</td>
<td>Unknown</td>
<td>No</td>
<td>Yes</td>
<td>Line-drawing</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Manual</td>
</tr>
<tr>
<td>Proposed System</td>
<td>Geometric</td>
<td>46/143</td>
<td>No</td>
<td>Yes</td>
<td>Photo-realistic</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Automatic</td>
</tr>
</tbody>
</table>
3.4 Conclusions

This chapter gave a review of different caricature generation systems, which included geometric approach, linguistic approach and commercial caricaturing machines. It first, introduced the ideas and implementations of the available caricature generation systems. It then proceeded to discuss the strengths and weaknesses of each of them. It further summarised the common problems of different caricature systems, where most of them ignore how professional caricaturists exaggerate non-linearly and apply different exaggeration factors for different facial components. As a result, a caricature generation system with the ability to capture the drawing style of a particular artist from his/her products is proposed, which is expected to overcome the common problems in the existing systems. Finally, a comprehensive table is presented to compare the characteristics of different caricature generation systems.

Learning from the products of an artist in order to find out the exaggeration rules embedded in his/her mind is a challenging task, which helps the people without inborn talent to learn how to draw caricatures by solving the mystery in caricaturing.
Chapter 4

Introduction and Review of Background Technologies

4.1 Introduction

Automatic caricature generation with embedded drawing style of an artist is a challenging research topic, which involves a combination of different technologies. It requires a generic 2D face modelling framework that can represent any human face as a platform for caricaturing. An artificial intelligence technique is then applied to accomplish the drawing style capturing task. Further image processing technologies, such as morphing [63] and warping [64], are needed to convert an input image into a caricature.

This chapter gives an introduction to and review of different background technologies that have been used within the research context presented in this Thesis. It is organised as follows:

4.2 What is a human face?
4.3 Review of 2D geometric face models
4.4 Introduction to artificial intelligence
4.5 Introduction to image processing technologies
4.6 Chapter summary and conclusions
4.2 What is a Human Face?

4.2.1 Introduction to Human Face

Human face is a complex biological structure. According to the Oxford Dictionary of English [21], the face is defined as:

"The front part of the head, from the forehead to the chin, the visage, countenance".

4.2.2 Facial Components

The following are the major facial components considered in human face modelling:

1. Hair
2. Forehead
3. Eyebrow
4. Eye
5. Ear
6. Cheek
7. Nose
8. Lip
9. Teeth
10. Tongue
11. Mouth
12. Chin
13. Neck

Figure 4-1. Facial components of a human face.
4.3 **Review of 2D Geometric Face Models**

In human face representation research, generic face models are commonly adopted to describe various human faces through parameter inputs from users [65-67]. Although face modelling studies commenced in the 1970s, the complexity of human facial appearance has resulted in face modelling being still considered as an open research problem. The applications of a human face model can be extended to a variety of areas, such as art, entertainment, business, sport, medicine, education, etc. As the requirement of a face representation increases, the accuracy and reality of face models become the major research concern.

Geometric face modelling research can be divided into three categories: Parametric models, physically based models and feature point based models. A detailed review of face modelling techniques can be found in the book by Parke and Waters [44].

**4.3.1 Parametric Models**

The first parametric model was introduced by Parke in 1972 [65-67]. The basic concept of parameterisation can be considered as a class of objects in which each member has its own distinctive differences. The differences between member objects are associated with a set of specification criteria, which are also referred to as parameter values. A complete (ideal) set of criteria allows specifying a member of an object class, which contains any possible facial expression of a human face, by choosing appropriate parameter values. However, developing a complete set of parameters is difficult even for a simple object.

Although defining the whole set of facial parameters is time consuming, Parke's face model is easy to use. It only requires a set of appropriate parameters to specify a face,
which provides the advantages of simple and efficient data handling.

4.3.2 Physically Based Models

A human face consists of skin, muscles and bones. All these fundamental elements are considered in a physically based face model. The groundwork, Facial Action Coding System (FACS), was laid by Ekman [68] in 1978, which was developed from analysing the anatomical basis of facial movement. It introduced the concept of Action Unit (AU) that simulates an action produced by one muscle or group of related facial muscles. Platt [69] adopted the AU concept and proposed the first physically based face model in 1981. All muscle movements of the model are based on FACS that consists of 38 regional muscle blocks interconnected by a spring network, which is then deformed by simulated muscle forces in order to generate AUs. Waters [70] further improved the model by using vector muscle to imitate different muscular actions upon skin. Afterwards, Terzopoulos and Waters [71] introduced different skin layers into the model to handle highly realistic and subtle facial movements.

In 1988, Magnenat-Thalmann [72-73] proposed another physically based face model that uses Abstract Muscle Actions (AMA) instead of FACS. AMA concentrates on how facial muscle movement procedures affect the appearance of a human face. The order of the procedures is essential as they are dependent on each other.

4.3.3 Feature Point Based Models

Feature Point Based Model is a face model that uses a set of landmarks to locate the most important facial features. A facial mesh can be constructed by connecting these feature points in a particular manner to represent a complicated human face [44]. Facial animation can also be created by simply shifting the feature points along a time
Chapter 4. Introduction and Review of Background Technologies

series. Since the first feature point based model introduced by Brennan in 1985 [2], different novel models have been proposed [4-8,12,49] or extended based on Brennan's design [3]. The numbers of feature points in these models are similar to each other, as they vary slightly from application to application.

Brennan [2] is a forerunner of research focused on the Feature Point Based Face Modelling. She simplified the implementation of a parameterised facial model, as discussed in 4.3.1, and developed a face model with 160 feature points in her caricature generation system. These feature points have been considered as the best locations to represent a face, and which are general for any human face. The advantage of using such a model is that the feature points can also be used to represent caricatures that have exaggerated human features and expressions. Besides, feature point based model is simple and flexible in its application. Any system with specific focus on a particular facial component can simply increase the number of feature points to it. However, due to automatic marking of feature points accurately is not an easy task, manual marking of feature points has become a common approach [2-3,11-14].

4.3.4 MPEG-4 Standard

In view of the above models, various approaches have been proposed during the last 30 years. However, different researchers adopt different models that cause their systems to be incomparable and incompatible with each other. Hence, a standard face model was found to be essential to unify researchers and image coders. As a result the Motion Picture Expert Group (MPEG) [49-51], which aims at standardising technologies and provides efficient image and video description, storage and transmission, formed a standard, MPEG-4, for human face modelling and animation.
in 1999. The full technical documentation of standardised face modelling approach is documented under ISO/IEC MPEG-4 Part 1 (System) [49] and Part 2 (Visual) [50].

A clearly defined face model was proposed within MPEG-4. It describes a human face, based on so-called Facial Definition Parameters (FDPs). FDPs are a set of facial parameters that consists of 84 feature points, as shown in figure 4-2. All the feature points can be used for face calibrations, and each of them is associated with a unique number. Feature points are categorised into different facial groups according to the region they belong to. Their locations are represented as coordinates in 2D, i.e., (x,y) or coordinates in 3D, i.e., (x,y,z). The specification of each feature point can be found in Appendix A.

![Figure 4-2. MPEG-4 facial definition parameters [49].](image-url)
4.3.5 Choosing a Geometric Face Model for 2D Caricaturing

The aim of this project is to capture the drawing style of a particular artist from his/her 2D caricature products. In order to choose an appropriate face model, the following requirements have to be considered. First of all, the face model must be in a 2D basis. Besides, as both input (original image) and output (generated caricature) to the final system are static 2D images, the process of deforming the input into the corresponding output is not a responsibility of the face model. As a result, no animation support is required from the face model. Moreover, simplicity and efficiency should be emphasised. It should be noted that the errors and handling time decrease when the simplicity and efficiency of the model increase.

In a parametric model, the development of a complete set of parameters is difficult and time consuming. Added to this, the resulting model is less realistic than the physically based model as the biological consideration of human head is not involved. Consequently, the use of parametric approaches faded out in 1980s.

A physically based model is a complicated model that focuses on how facial muscle movements affect facial expressions, which provides smooth and realistic facial expression changes during animation. However, face animation is not a concern of this project and the use of physically based models is not solicited.

In feature point based models, the idea of locating facial components by feature points is an approach that can be effectively used in 2D face caricaturing. Firstly, marking feature points on an image is straightforward. Although automatic marking of facial components precisely is difficult, manual marking can be adopted instead within an experimental context. Manually marking feature points allows applying the face
model on images accurately in a reasonable period of time. Secondly, feature point based models fully support both 2D and 3D systems, which provide high flexibility in both system design and extension.

Brennan proposed the first feature point based model for 2D caricaturing [2], and it has since become a benchmark algorithm for continuing research [3-8]. Nevertheless, since the launch of the MPEG-4 standard in 1999, human face modelling and caricaturing research [12,98-101] tended to follow it. MPEG-4 standard is widely expected to be setting the trend in face modelling and caricaturing research in the near and distance future. Further, for practical reasons, adopting old models is not recommended once a new authoritative standard has been developed.

In view of the above, the most appropriate face modelling approach for this project is the feature point based model, adopted by the MPEG-4 standard. As a result, the feature point based model proposed in the MPEG-4 standard is chosen as the basis of the project. Although the FDPs of the model may not perfectly fit the required system design, modifications can be made based on the provided feature points. This will be further discussed in Chapter 5.
4.4 Introduction to Artificial Intelligence

4.4.1 An Overview of AI

Artificial intelligence (AI) had an explosion of renewed interest in the past two decades, though it originated approximately fifty years ago. A significant amount of expectations have been based on innovations of AI technology in the coming decades, despite the fact that the potential of development and application of AI technologies are somewhat unpredictable.

4.4.2 The Definition of AI

Artificial intelligence is defined as “The study of ideas which enable computers to be intelligent” by Winston in 1984 [74]. Patterson further gave a detailed definition of AI in 1990 [75]. “AI is a branch of computer science concerned with the study and creation of computer systems that exhibit some form of intelligence: systems that learn new concepts and tasks, systems that can reason and draw useful conclusions about the world around us, systems that can understand a natural language or perceive and comprehend a visual scene, and systems that perform other types of feats that require human types of intelligence”.

4.4.3 Different AI Technologies

There is no doubt that artificial intelligence has been widely used in various areas such as medicine, defense, surveillance, economics, banking, chemistry, etc [75]. However, different AI technologies, including artificial neural networks, case based reasoning, data mining, fuzzy logic, genetic algorithms and knowledge based system, have their specific strengths and characteristics. Therefore careful selection of the most appropriate AI technology is one of the keys to success in its application.
4.4.4 Artificial Neural Networks

An Artificial Neural Network (ANN), commonly referred to as a "neural network", was introduced by McCulloch and Pitts in 1943 [76]. It is an effective machine learning technology that attempts to imitate the way a human brain works. The capacity of learning from experience results in a system that can continuously self-improve and increase effectiveness. Common applications include, data mining systems that try to discover rules from data sets and information filtering systems that automatically capture the interests of users [77].

(i) Biological Neuron

There are over 10 billion neurons in a human brain, which are the basic units that provide abilities of thinking, remembering, and experiencing sensations to us. All these neurons communicate with each other through 60 trillion connections (synapses). Each neuron has input channels (dendrites) that receive biological impulses from other neurons. The cell body combines the inputs in a predetermined way and performs a function to modify the combination. Finally, the result is transmitted as an output to other neurons by the axon [77].

![A simplified biological neuron](image-url)
(ii) **Artificial Neuron**

Artificial neurons are the basic units in an artificial neural network that try to simulate the function of biological neurons. An artificial neuron receives one or more input signals and subsequently multiplies each input signal by a corresponding embedded weight and adds the results together. To obtain the output signal, it further applies a non-linear activation function. These output signals are then passed to other neurons within the network, where the same process is repeated [77].

![Figure 4-4. A schematic representation of an artificial neuron [78].](image)

(iii) **Neural Network Architecture**

There are two main categories of neural network structures: feed-forward networks and recurrent networks. A feed-forward network is a more common architecture that represents a function of its current input, where no internal states other than the weights themselves are considered. On the other hand, a recurrent network feeds its
output back into its own inputs and forms a bi-directional relationship. The response of the network to a given input depends on its initial state or previous inputs, which imitates short-term memory in a human brain [77]. This section will concentrate on the feed-forward network as its simplicity made it a more popular choice for the research presented in this Thesis. Further, the ability of short-term memory is not required in this project.

(iv) Neural Network Layers

Feed-forward networks are arranged in layers, where each neuron receives input only from neurons of the previous layer. There are two sub-categories: single layer networks and multi-layer networks.

A single layer feed-forward network, which is also known as a perceptron network, is a neural network in which all inputs connect directly to the outputs, without any hidden layer. Conversely, a multi-layer network has one or more hidden layers, which benefits from increasing the functional complexity that the network can represent. Although there is no standard rule to decide the number of hidden layers in a neural network, the trial and error method is widely recognised as one of the best solutions to find out the most effective number of hidden layers [77]. Masters [79] further provided some guidelines to neural network researchers: “With a single, sufficiently large hidden layer, it is possible to represent any continuous function of the inputs with arbitrary accuracy. With two layers, even discontinuous functions can be represented”. He also stated that “The required number of artificial neurons in the hidden layer is not immediately apparent, but is normally near the square root of the product of the number of input nodes multiplied by the number of output nodes.” [79]
(v) Training a Neural Network

When the neural network architecture has been decided, the next imperative step is training. A neural network can be trained in either supervised or unsupervised way. Supervised training is the most common training method, which requires samples for the neural network to learn from. The collective name of all the samples for a particular training is referred as the training set. Firstly, a user feeds an input into the neural network and specifies the desired output. Then the calculated output from the network is compared to the output that has been specified by the user. If the output is undesired, the connection weights embedded in the neurons of the network are modified until the network functions accurately. By using this method, the errors in the network results can be minimised gradually [78].

The most common supervised learning algorithm in training feed-forward networks is back-propagation, it was popularised by Rumelhart and McClelland in 1986 [80]. The algorithm first computes the error of the output by using a delta rule, which is also named as the gradient-descent method [81], to calculate the difference between the output vector and the correct answer provided by the training set. No learning will occur if the difference is zero; otherwise, a weight change value will be calculated and propagated backwards through the network to update the weights of each layer. The whole process is repeated until all training instances have been processed.

Another way to train the neural network is unsupervised training. Similarly to the supervised training, a training set with sample inputs are fed into the network, but no desired outputs are provided. The training process allows the network to organise its hidden neurons and find out the distinctive features of the inputs. The input data must be carefully prepared so that the unique feature of an instance can be discriminated by
the neural network [77].

(vi) Neural Network Validations

The accuracy and capability of a trained neural network should be evaluated and scrutinised before putting it into practice, as small errors could result in a misleading output [78]. The evaluation process is known as validation, which reveals the reliability of a trained neural network.

A validation set is prepared by collecting a set of samples other than those in the training set, which is then fed into the network for evaluation. The outputs provide an unbiased assessment of the network when comparing with the correct answers. If the result of the network is undesired, it could be due to an insufficient or incomplete training set. The network can also suffer from over-training if it has learnt not only the basic mappings between inputs and outputs, but also memorised the subtleties and errors specific to the training set. This causes the network to be too dependent on the cases studied and therefore reduces the generalisation ability to the new data [77]. As a result, retraining of the neural network is required until satisfactory result is obtained.

(vii) Capabilities of Neural Networks

The major capabilities of neural networks were summarised by Haykin in 1990 [82]:

1. Non-linearity. A neural network is made up of neurons with embedded linear or non-linear equations, which are able to compute and provide results non-linearly.

2. Input-output mapping. In neural network training, the inputs and outputs of the training set are mapped automatically. Hence, a detail understanding of the
mapping process is not necessary for users.

3. Adaptivity. Neural networks have an effective capacity of changing neuron weights according to the environment, which can also be designed to adapt in real time. Any neural network trained to deal with a specific environment can be easily retrained to adapt to different operating environmental conditions.
4.5 Introduction to Image Processing Technologies

4.5.1 An Overview of Digital Images

A digital image is a rectangular matrix that consists of individual pixels (short for Picture Element) arranged in rows and columns. The pixel is the basic element of a digital image and represents a single spatial location of it. Each pixel uses a bit-depth value to represent its colour. Meanwhile, the number of bits of the bit-depth value determines how many different colours of the pixel can be displayed. The collective visual effect of all closely arranged colour pixels creates an image [83].

4.5.2 Basic Image Processing Technologies

Digital image processing is a technique that manipulates information on both input and output digital images, which includes photographs and video frames. Most image processing technologies consider a digital image as a 2D signal and then apply signal processing techniques on it [83]. As digital image processing is an extensive topic, only techniques relevant to the project are covered in this section. Applications of these techniques within the current research can be found in subsection 4.5.3.

(i) Nearest Neighbour Interpolation

Nearest Neighbour is the simplest interpolation method that is commonly used in digital image resizing. It calculates the closest corresponding pixel from the source image to represent each pixel in the destination image. However, the resulting image usually suffers from aliasing effects where jagged edges heavily exist, as no colour averaging is performed. Nearest neighbour only works fine if the digital image is rectangular in content, and the enlarge/reduce scale is a multiple of two (e.g. 2x, 4x, 6x...) [84].
(ii) Bilinear Interpolation

An improved approach, known as bilinear interpolation, is proposed to tackle the aliasing problem. It first locates the closest corresponding position from the source image for each pixel in the destination image by using the nearest neighbour method mentioned above. Subsequently it calculates the colour of the pixel in the destination image by averaging 4 pixels (top, bottom, left and right) that surround the located pixel on the source image. After finishing all pixel calculations in the target image, an anti-aliasing final result without jagged edges can be obtained as each pixel was obtained by a weighted combination of the neighbouring pixels that result in a progressive change of colour [84]. Figure 4-5 illustrates an example of reducing the size of an image using bilinear interpolation.

Figure 4-5. Reducing the size of an image by bilinear interpolation [85].
(iii) **Bicubic Interpolation**

Bicubic interpolation is the most sophisticated approach that provides the best quality output, which further improves bilinear interpolation by sampling data in two dimensions. It calculates the colour of the pixel in the destination image by adding 8 pixels (top, bottom, left, right, top left, top right, bottom left and bottom right) that surround the located pixel in the source image, which means all adjacent pixels from any directions are considered. The sum is then averaged with preferences based on a pre-defined weighting function. The final result is the best amongst the three methods described above, though the computational requirement for bicubic interpolation is the highest [84].

![Figure 4-6. (a) Nearest neighbour. (b) Bilinear interpolation. (c) Bicubic interpolation.](image)

[Note that nearest neighbour simply replicates the number of pixels. Bilinear interpolation samples from both vertical and horizontal directions. Bicubic interpolation samples from vertical, horizontal and diagonal directions] [86]

### 4.5.3 Applications of Image Processing Technologies

The technologies briefly discussed above form the basis of common image processing applications such as, image scaling, rotating, warping and morphing. As these applications, commonly available in most desktop photo-editing software, are popularly used in automatic caricaturing, their fundamentals are briefly explained below:
(i) Image Scaling

Image scaling is a geometric transformation technique of image processing, which handles size enlargement or reduction of digital images. Three different techniques mentioned above are capable of accomplishing the task, while the algorithm selection is just a tradeoff between speed and quality. Nevertheless, bicubic interpolation is the most common method for image scaling as it provides the best anti-aliasing results [84].

(ii) Image Rotation

Image rotation is another common geometric transformation technique. Although a number of algorithms have been proposed, the underlying ideas are more or less the same. One of the common approaches is very similar to image scaling, which first applies the nearest neighbour algorithm, with a rotation angle involved in calculation, to calculate the closest corresponding pixel from the source image for each pixel in the destination image. Afterwards, the final output image is computed using bilinear interpolation. The result is normally larger than the original image as extra regions are added to represent pixels that do not exist in the original image [84].

Figure 4-7. An example of image rotation.
(iii) Image Warping

Image warping is another geometric transformation technique that deforms the content of a digital image, according to the control points provided. The algorithm was proposed by George Wolberg in 1990 [63], since then it has been widely applied in computer graphics and film productions to create funny human faces with distortions [52]. Most picture editing software packages [87-88] provide robust warping functions and user-friendly environments to users, which allow digital image warping to be done in just a few clicks. A warping program accepts an original image, a source mesh and a target mesh from the user. The source mesh specifies the coordinates of control points in the source image, while the target mesh specifies their corresponding shifted positions in the output image. Both meshes must have the same dimensions in order to establish a one-to-one correspondence. The warping program then deforms the original image towards the target image based on the information from both meshes. Figure 4-8 demonstrated a simple example of digital image warping.

<table>
<thead>
<tr>
<th>Input:</th>
<th>Original image</th>
<th>Source mesh</th>
<th>Target mesh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image1" alt="Original image" /></td>
<td><img src="image2" alt="Source mesh" /></td>
<td><img src="image3" alt="Target mesh" /></td>
</tr>
</tbody>
</table>

Output:

Target image

Figure 4-8. A simple example of digital image warping.
Chapter 4. Introduction and Review of Background Technologies

The algorithm of warping is basically an unequal resizing of different regions, specified by meshes, within an image. When comparing the source mesh and target mesh in figure 4-8, the size of the top row in the target mesh remains unchanged. However, the size of the middle row is enlarged while the bottom row is reduced. These result in the target image with an elongated face whereas the neck is shortened, which demonstrated the effects of partial enlargement and partial reduction within an image. The resizing algorithm can be accomplished by either bilinear or bicubic interpolation [64].

(iv) Image Morphing

Image morphing, also known as a transition morphing method, provides a smooth animation effect to illustrate how an image is deformed into another image. The morphing program accepts an original image, a target image, a source mesh and a target mesh. The idea was built on top of warping, with an intermediate image of the original and the target images constructed [63]. A simple example explains the details in figure 4-9.

![Figure 4-9. An example of average mesh calculation.](image)

In figure 4-9, the outermost and the innermost squares represent the source and the target meshes respectively, they are overlapped together with the central point as the reference. In order to morph from the source square into the target square, the system
first calculates an average mesh according to the control points (i.e. corners) of both meshes, which is represented by the middle square in figure 4-9. After that, the morphing program reduces the size of the source image to the size of the average mesh, and in the meantime enlarges the size of the target image to the size of the average mesh. Once the source and target images have exactly the same size and overlap with each other, an intermediate image can be generated by averaging the colour of each corresponding pixel. The interpolation used above can either be bilinear or bicubic. Finally, a dissolving algorithm is applied to produce a visual effect that demonstrates how the source image is animated to the intermediate image, and from the intermediate image to the target image. For more information about morphing, the reader is referred to a more sophisticated approach presented in “Feature-based Image Metamorphosis” [52].
4.6 Chapter Summary and Conclusions

The aim of this chapter was to introduce and review different background technologies adopted within the context of the main research focus of this Thesis. It first provided a review of research in 2D geometric face modelling, which focused on the available models and their underlying technologies. It provided a definition of a human face and introduced different facial components. It further discussed the advantages and disadvantages of the existing 2D geometric face models. MPEG-4 face modelling framework and related Facial Definition Parameters (FDPs) were discussed in detail and was justified to be the most suitable approach to face modelling within the novel automatic caricaturing approaches to be presented in Chapters 6 and 7. The chapter proceeded to give an overview of artificial intelligence and which was followed by an in-depth discussion of the concepts, structures, capabilities and usage of artificial neural networks. Finally, the chapter focused on a discussion on basic image processing technologies to be used within the proposed automatic caricaturing techniques of Chapters 6 and 7. It introduced the algorithms of nearest neighbour interpolation, bilinear interpolation, bicubic interpolation and their common applications in image scaling, rotating, warping and morphing.

The MPEG-4 standard provides an excellent framework for geometric face modelling. However the comprehensive set of FDPs defined within the MPEG-4 standard has been included with applications in facial animation in view. Consequently, for caricaturing applications a smaller but revised set of MPEG-4 FDPs can be used. Chapter 5 focuses a novel design of a face modelling framework for automatic caricaturing applications, based on the MPEG-4 standard's face modelling framework.
Neural Networks can be effectively used to incorporate an effective learning ability to computer based systems. Although the thinking and reasoning processes in human brains are too difficult to be completely imitated by any existing artificial intelligence technology, the state-of-the-art neural network techniques are able to address this issue to a significant extent. It has been shown that caricature drawing depends on a complicated thinking process in an artist’s mind (see subsection 3.3.3). Therefore, in order to capture the drawing style of an artist, a neural network can be used to provide a promising solution. In Chapters 6 and 7 it is shown that a neural network is able to discover the drawing rules based on a particular artist’s products and incorporate them within a computer based system. The resulting system is further shown to be capable of conveying the captured drawing style via caricature generations, with the aid of image processing technologies.
Part Two

Contributions of Research
Chapter 5

Proposed Face Modelling
Framework and Dataset Preparation

5.1 Introduction

The pros and cons of existing 2D geometric face models have been fully discussed in Chapter 4. A novel face modelling framework with two geometric face models is proposed in this chapter. After defining the novel face models suitable dataset preparation that includes facial image selection, caricature drawing, normalisation, manual marking of feature points and mean face generation, are discussed. The proposed face models serve as platforms for the caricature generation approaches that are proposed in Chapters 6 and 7.

The chapter is organised as follows:

5.2 A novel geometric face model with 46 feature points
5.3 An enhanced geometric face model with 143 feature points
5.4 Comparing proposed face models with MPEG-4
5.5 Dataset preparation
5.6 Chapter summary and conclusions
5.2 A Novel Geometric Face Model with 46 Feature Points

As a result of the critical review of existing geometric face models in Chapter 4, the Facial Definition Parameters (FDPs) of MPEG-4 standard [49] was selected as the basis for feature point selection and facial representation within the context of the proposed research. The reasons for this choice were explained in Chapter 4. However, as the facial feature points defined within the MPEG-4 standard were aimed at facial animation, they are not entirely suitable for caricaturing purposes. Typical caricaturing systems require less accuracy in high movement facial components (e.g., lips) and more accuracy in facial components having highly varying and unique shape characteristics (e.g., eyebrows). Therefore in this chapter several adjustments are proposed to the set of MPEG-4 FDPs for framework’s adaptation in caricaturing applications. The MPEG-4 FDP framework provides the basis of the proposed 46 feature point based face modelling framework detailed below.

The Facial Definition Parameters (FDPs) of MPEG-4 standard that has been introduced in Chapter 4 is revisited in figure 5-1 as a reference for the following sections.

![Figure 5-1. MPEG-4 facial definition parameters [49].](image_url)
5.2.1 Eyebrows

The MPEG-4 standard uses three FDPs to outline the contour of each eyebrow (see figure 5-1). However due to the detailed attention often given to the shape exaggerations of eyebrows in caricaturing and the adverse nature of shape of human eyebrows, preliminary observations by the author concluded that three FDPs are insufficient for defining the eyebrows of human face. Therefore the inclusion of a new FDP at the boundary of each eyebrow is proposed for increasing the representation accuracy in caricaturing. The new FDPs are marked as 4.7 and 4.8 on the left and right eyebrows respectively, which are vertically below 4.3 and 4.4 to define the thickness of the eyebrows, as illustrated in figure 5-2.

5.2.2 Hair

According to the MPEG-4 standard, only three FDPs (11.1, 11.4 and 11.5 of figure 5-1) are defined for the representation of hair and hairline. These FDPs are obviously not sufficient for accurately representing most hairstyles. Therefore it is clearly evident that the MPEG-4 standard does not concentrate on the subject of hair modelling within its scope. More realistic geometric hair modelling techniques have been proposed as standalone research [89-90], with most frameworks using over hundred feature points due to the need of addressing the complex nature of uncountable hairstyles. Unfortunately the accurate modelling of hair within the context of current research becomes a memory intensive, tedious task. Therefore for simplicity, the accurate caricaturing of hair has been ignored within the present design.
5.2.3 Mouth

The definition of FDPs within the MPEG-4 standard is mainly driven by the need of a smoother, more realistic facial animation applications. Therefore a large number of feature points have been defined to accurately represent high motion facial components such as lips. Such representation accuracy is not necessary for lips in an automatic caricaturing system. As a result, only feature points at the corners of the mouth that define the mouth contour (see figure 5-3) can be considered in caricaturing.

5.2.4 Eyes

Similarly, the large number of feature points defined for accurately animating eye blinking within the MPEG-4 standard are not required in caricaturing applications. Consequently, the feature points describing the movement of eyelids, 3.2, 3.4, 3.1 and 3.3 (see figure 5-1), are discarded.
The feature points of irises are also removed after being used for normalisation (see subsection 5.5.3).

5.2.5 Nose

Likewise, some feature points of the nose can be removed. For instances, the feature points 9.12 and 9.3 of figure 5-1 that represent the tip of the nose are only meaningful in a 3D but not in 2D face model. Therefore they are ignored in the proposed model. Moreover, the feature points, 9.13, 9.14 and 9.15 (see figure 5-1), that describe the internal details of the nose are also removed for simplicity. Only those feature points that describe the basic outline of the nose, i.e. 9.6, 9.7, 9.2, 9.1, 9.4 and 9.5, are preserved.

5.2.6 Other FDPs

Further within the present research context of the project, some subtle and static facial feature points, such as those representing the cheeks and dimples, are excluded from consideration. Furthermore the FDPs defined for the neck have been entirely ignored as this research focus is only facial caricaturing whereas all FDPs representing ears have been preserved.

As a result of the reduction of the FDPs in the proposed framework, the required computer resources during the drawing style capturing process can be maintained at a manageably low level. Note that the reduction of FDPs as defined in the original MPEG-4 standard is not essential in caricaturing applications provided unlimited
processing power and memory space were available. However in practical applications where resource limitations are a bottleneck, a compromise of the number of FDPs is highly solicited [2,14].

5.2.7 Overall

The facial feature point modifications described above form a novel geometric face model with 46 feature points, which is used in the drawing style capturing approach proposed in Chapter 6. This face model is named as the “Simple Face Model” (SFM) and is fully illustrated in figure 5-6.

![Figure 5-6. The proposed Simple Face Model (SFM).](image-url)
5.3 An Enhanced Geometric Face Model with 143 Feature Points

Similar to that of the Simple Face Model, the FDPs of MPEG-4 standard provides the basis for the FDPs defined within the second proposed face model. Within this framework rather than an overall reduction of the number of FDPs as compared to the standard MPEG-4 FDPs, a number of additional FDPs are added to different facial components for increasing their representation accuracy. Note that this model is used for the facial component-based caricature generation of Chapter 7, where the neural network based capture of the drawing style is separately carried out for individual facial components. As a result the excessive computational resource requirement does not provide a significant bottleneck, as it would have caused if the framework was used in association with the entire face-based automatic caricaturing approach. The following sections explain the modifications proposed for the FDP set of each facial component:

5.3.1 Eyebrows

In the Simple Face Model, four feature points are used to outline the contour of each eyebrow whereas only three were used in the MPEG-4 framework. In order to provide a more accurate representation of eyebrows, 16 feature points are proposed to each eyebrow, which are expected to be capable of describing complex curves that define typical human eyebrows and more importantly, eyebrows of freehand caricatures. Note that the newly included FDPs are distributed evenly around the eyebrows of the proposed model and are left unnumbered in figure 5-7, for clarity.

Figure 5-7. Modified eyebrows FDPs.
Chapter 5. Proposed Face Modelling Framework and Dataset Preparation

5.3.2 Hair

Similar to that of the Simple Face Model, the caricaturing of the hair has been ignored for simplicity, within the current model. Consequently, the feature point, 11.4 (see figure 5-1), in the MPEG-4 face model that represents the hairline has been discarded (see subsection 5.2.2).

5.3.3 Mouth

In the Simple Face Model only 5 feature points were used to represent the mouth. In the current face model a further 11 FDPs are proposed to be added to defining the mouth with the aim of improving the accuracy of its representation in caricaturing. To this effect 5 new FDPs are added to the upper lip while 6 new FDPs are added to the lower lip. The newly added points are evenly distributed around the mouth as illustrated in figure 5-8.

5.3.4 Eyes

Once again, feature points of the eyes that are not describing their contours are discarded, as the shapes of the eyes are the main concern in caricaturing (see subsection 5.2.4). Despite this, 8 extra feature points are added to each eye in the current framework, 4 to the upper eyelid and another 4 to the lower. The proposed feature
points are able to provide a more detailed representation of different eye shapes, which usually various from oval shape to circular shape.

5.3.5 Nose

Further to the feature points proposed in the Simple Face Model, an additional 16 feature points are added to the outline of the nose. According to observations, the variations of nose shapes are insignificant in the original facial images, but significantly large in the caricatures drawn by artists. Therefore to capture the diverse contours of the nose, more feature points are recommended to strengthen the face model. Note that only two FDPs are added to represent the bottom edge of the nose, all the rest are distributed evenly along both sides of the nose.

5.3.6 Ears

The FDPs of ears outline their shapes. To facilitate more accurate ear representations, 6 further feature points are added to each ear of the Simple Face Model, which result in 11 feature points per ear in total in the current face model. These additional feature points are mainly located at the outside edges of the ears, as shape changes are more intense than in the inner part.
5.3.7 Face Outline

The outline (perimeter) of a face is one of the features that are frequently exaggerated by caricaturists, as modifying the contour of the face could easily produce humorous effects [14]. Though the exaggerations to the human body is not considered within the research context of this project, an exaggerated face with a small body is a common technique that is adopted by most artists. As a result, 27 feature points are used in the proposed model to improve face representation. These feature points are distributed along the face contour from the point defining the hairline, 11.1, to the chin, 2.1, with 12 points on each side. The proposed feature points are expected to be able to accurately represent most kinds of facial exaggerations performed by caricaturists, especially the highly modified jaw and chin areas.

5.3.8 Other FDPs

Similar to the Simple Face Model proposed in section 5.2, the subtle facial feature points that represent the cheeks and dimples are removed for simplicity (see subsection 5.2.6). It should be noted that although a significant number of additional feature points are added to the current face model, the additions are only aimed at improved representation accuracy of facial components/features that are important in caricaturing applications.
5.3.9 Overall

All facial feature point modifications described under section 5.3 define a novel geometric face model with 143 feature points, which is adopted by the facial component-based drawing style capturing approach proposed in Chapter 7. This final face model, illustrated in figure 5-13, will be here forth called as the "Enhanced Face Model" (EFM).

Figure 5-13. The proposed Enhanced Face Model (EFM).
5.4 Comparing Proposed Face Models with MPEG-4

In table 5-1 the number of FDPs defined in the MPEG-4 standard [49] is compared against the number of FDPs proposed in the two proposed face models, on a facial component-by-component basis. Further in figure 5-14, the FDPs of MPEG-4 standard and the FDPs of two proposed models are illustrated. Note that for clarity of representation, some facial feature points are left unnumbered.

<table>
<thead>
<tr>
<th>Region</th>
<th>No. of FDPs in MPEG-4</th>
<th>No. of FDPs in the Simple Model</th>
<th>No. of FDPs in the Enhanced Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td>15</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>Left Ear</td>
<td>5</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Right Ear</td>
<td>5</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Left Eyebrow</td>
<td>3</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>Right Eyebrow</td>
<td>3</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>Left Eye</td>
<td>7</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Right Eye</td>
<td>7</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Nose</td>
<td>11</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Mouth</td>
<td>18</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>74</td>
<td>46</td>
<td>143</td>
</tr>
</tbody>
</table>

Table 5-1. Comparison of the number of FDPs of MPEG-4 and two proposed models.

(a) MPEG-4 Model (74 pts). (b) Simple Face Model (46 pts). (c) Enhanced Face Model (143 pts).

Figure 5-14. Comparison of MPEG-4 and proposed face models.
Chapter 5. Proposed Face Modelling Framework and Dataset Preparation

5.5 Dataset Preparation

5.5.1 Face Selections

Various human face databases [91-95] with large number of facial images are available to researchers worldwide for testing / comparing algorithms and for benchmarking purposes. While different databases have their own characteristics, choosing an appropriate database for an application should consider the image quality, sizes, lighting conditions, and object postures.

The Purdue University, USA, online Aleix Martinez and Robert Benavente (AR) face database [96] was chosen to provide original facial images for the research presented in this Thesis. This database consists of a set of high resolution facial images of size of 256 x 256 pixels each, captured under controlled conditions. The clarity of the facial images has made the AR database a popular choice in research, its use in the proposed work further enables researchers to easily compare this research with existing work. Further the fixed pose maintained in capturing the facial images of the database, greatly decreases the complexity of normalisation step that will be detailed in subsection 5.5.3.

The dataset of this research has been further limited to a collection of male facial images with short hair and no facial accessories, such as glasses, makeup, hats, hairpieces and jewellery. These restrictions contribute to the simplicity of the experimentation, which helps focus the proposed research around drawing style capture based automatic caricature generation, without being drawn to non-facial component based changes that have the potential in misinterpreting a caricaturist’s drawing style.
A total of twelve images from the AR face database were selected and subsequently forwarded to the next stage, i.e. inviting artists to produce caricature drawings.

5.5.2 Inviting Artists to Draw Caricatures

Three professional caricaturists from different countries were invited to produce caricature drawings. These three artists are here forth named as the first, second and the third artists.

Each artist was invited to draw a caricature for each of the twelve original images. In order to fully explore the drawing styles of the artists, no drawing constraints were forced upon them, i.e., they were allowed to exaggerate or distort any facial component according to wishes, in contrary to the method adopted in Liang et al.’s system [15].

However finding and getting agreement from caricaturists to be involved in the production of a large number of caricatures has been a major challenge continuously faced by the proposed research. Further it is unreasonable to request an artist to draw a large number of caricatures to enable critical analysis. As a result, the key additional challenge of the research presented in this Thesis is to capture the drawing style of an artist based on a limited dataset and then generate high quality caricatures that embed the artist’s drawing style. Further practical challenges will be discussed in Chapter 8.
5.5.3 Normalisations

Although the original images and the caricatures are captured under similar pose and scale, they are normalised to further convert them into the same scale and inclination level, so that the resulting images will be accurately comparable with each other. In the proposed work the normalisation approach proposed by Susan E. Brennan [2] was adopted, which assumes that the distance between two irises is a constant for all people. Even though the assumption may not be true for all human faces and caricatures, iris separation has the most advantages over any other reference points on the face in terms of reliability [97]. As a result, this normalisation approach has been widely used in previous research in automatic caricature generation and psychology [2-8, 12,14,31-34]. The advantages of choosing iris separation as the scaling factor are that it can be measured accurately and is independent of facial expressions as compared to other possible facial measures.

Once the XY coordinates of both irises are recorded, the distance between them, d, is given by

\[ d = \sqrt{(X_l - X_r)^2 + (Y_l - Y_r)^2} \]  \hspace{1cm} (1)

where l and r are corresponds to the left and right irises respectively. The scaling factor, \( \lambda \), is defined as

\[ \lambda = \frac{d_c}{d} \]  \hspace{1cm} (2)

where \( d_c \) is a constant value of irises distance. Moreover, the locations of irises can be used to determine the inclination level of a face, \( \Theta \) that can be used as the rotation factor. The formula is given by
The scaling factor is used to scale the face and the rotation factor is used to rotate the face. It is noted here that the methodologies used for image scaling and rotation have been presented in Chapter 4. The final normalised original image-caricature pairs obtained after normalisation are illustrated in figure 5-15.

\[ \theta = \tan^{-1} \left( \frac{Y_t - Y_r}{X_t - X_r} \right) \]
<table>
<thead>
<tr>
<th>Original Image</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Artist</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Artist</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; Artist</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>5</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>6</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td>7</td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td>8</td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
</tbody>
</table>
### Figure 5-15. The final normalised dataset.

<table>
<thead>
<tr>
<th>Original Image</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Artist</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Artist</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; Artist</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Original Image 9" /></td>
<td><img src="image2" alt="1&lt;sup&gt;st&lt;/sup&gt; Artist 9" /></td>
<td><img src="image3" alt="2&lt;sup&gt;nd&lt;/sup&gt; Artist 9" /></td>
<td><img src="image4" alt="3&lt;sup&gt;rd&lt;/sup&gt; Artist 9" /></td>
</tr>
<tr>
<td><img src="image1" alt="Original Image 10" /></td>
<td><img src="image2" alt="1&lt;sup&gt;st&lt;/sup&gt; Artist 10" /></td>
<td><img src="image3" alt="2&lt;sup&gt;nd&lt;/sup&gt; Artist 10" /></td>
<td><img src="image4" alt="3&lt;sup&gt;rd&lt;/sup&gt; Artist 10" /></td>
</tr>
<tr>
<td><img src="image1" alt="Original Image 11" /></td>
<td><img src="image2" alt="1&lt;sup&gt;st&lt;/sup&gt; Artist 11" /></td>
<td><img src="image3" alt="2&lt;sup&gt;nd&lt;/sup&gt; Artist 11" /></td>
<td><img src="image4" alt="3&lt;sup&gt;rd&lt;/sup&gt; Artist 11" /></td>
</tr>
<tr>
<td><img src="image1" alt="Original Image 12" /></td>
<td><img src="image2" alt="1&lt;sup&gt;st&lt;/sup&gt; Artist 12" /></td>
<td><img src="image3" alt="2&lt;sup&gt;nd&lt;/sup&gt; Artist 12" /></td>
<td><img src="image4" alt="3&lt;sup&gt;rd&lt;/sup&gt; Artist 12" /></td>
</tr>
</tbody>
</table>

Column 1 – Original images. Column 2, 3, 4 – Corresponding caricatures drawn by the first, second and third artists, respectively.
5.5.4 Manual Marking of Feature Points

A manual marking strategy is adopted for the selection of FDPs of the normalised original and caricature images. The advantages of this manual approach have been discussed in Chapter 4. Preliminary investigations by the author have revealed that under the context of this research, which focuses on proving that the proposed approach is capable of capturing the drawing style of an artist and is then able to automatically generate caricatures of the artist, this is a decision that provides reasonable accuracy and maintains simplicity. However many algorithms have been proposed in literature that are capable of automatic feature point marking [98-102] and are potential candidates to completely automating the feature point selection process of the proposed facial models. However it is noted that the accuracy and the reliability of these algorithms have to be critically evaluated before any attempts are taken to incorporate them into the proposed drawing style capture algorithm, as any inaccurate markings could lead to errors.

Fig. 5-16. Original (left) and corresponding caricature (right) with marked FDPs.

The software 'Morpher [103]' was used as a platform for feature point markings, which will be further discussed in subsection 5.5.5. Figure 5-16 illustrates an example of original image-caricature pair with marked FDPs. The (x,y) coordinate pairs of all manually marked feature points of the original face and the caricature, are then recorded and forwarded to the mean face generation stage.
5.5.5 Mean Face Generation

"Mean Face" refers to an average of faces one comes across during one's lifetime [31]. The concept of mean face has been fully discussed in Chapter 3. In order to implement this concept into the proposed drawing style capturing system, the mean face is obtained by averaging the normalised original images of the training set (i.e. images 1 to 10 in figure 5-15), using the 'Morpher [103]' software. Original images 11 and 12 (see figure 5-15) are reserved as the validation set. After manually marking the feature points by using the approach described in subsection 5.5.4, the (x,y) coordinate pairs of these points can be easily captured by the 'Morpher' software package, which subsequently generates the mean of two input images (at a given time) using image morphing technique. Note that due to the limitation of 'Morpher' software to only use two input images at a given time, the averaging of the ten facial images has to be done, pair-wise. This pair-wise approach is a common technique adopted by human face researches [104-106], which is able to effectively reduce the number of images by half in each generation level, and eventually obtains the final mean face under a binary tree scheme [104]. A simple morphing example that averages two faces using 'Morpher', is illustrated in figure 5-17.

Figure 5-17. An example of average face. (a) Image 1. (b) Image 2. (c) Average face of image 1 and 2.
In calculating the average of two facial images, the manually marked feature points play an important role. All the facial components, such as the mouth, nose, etc are first defined by the use of their FDPs. Subsequently, for each component, the feature points together are connected and the area enclosed is normalised, so that an area with equal texture is defined by each contour. A snapshot of the software ‘Morpher’ with feature points and referenced lines marked, is illustrated in figure 5-18.

Subsequently, the ‘Morpher’ averages the coordinates of the corresponding feature points and the colour values for each pixel. The underlying algorithm is known as “Image Morphing” [63] that has been discussed in detail in Chapter 4. Figure 5-19 illustrates the final mean face generated from the ten original images.
5.5.6 Summary

Figure 5-20, summarises all the dataset preparation steps described above.

![Diagram](https://via.placeholder.com/150)

Figure 5-20. Dataset preparation steps.

The recorded (x,y) coordinates of all feature points are forwarded to the next module, drawing style capturing algorithm, proposed in Chapter 6 for further analysis.
5.6 Chapter Summary and Conclusions

The chapter proposed a novel face modelling framework that is supported by two novel geometric face models. The two face models are developed based on the facial feature points defined by the MPEG-4 standard, with suitable modifications and extensions to cater for their specific use in caricaturing applications. The Simple Face Model with 46 feature points was defined considering computational and operational simplicity as the key design criteria. In Chapter 6 this model is adopted to demonstrate the efficiency of the entire face-based drawing style capturing algorithm. The second face model, Enhanced Face Model, increased the number of feature points to 143, with the aim of generating high quality caricatures when used in conjunction with the facial component-based drawing style capture algorithm to be presented in Chapter 7.

After proposing the face models, the chapter provided discussions on dataset preparation steps, which included selecting original faces, inviting professional artists to draw caricatures, normalisation, mean face generation and manual marking of feature points. The (x,y) coordinates of the marked feature points were recorded for further analysis in the forthcoming chapters.

Once all input data is readily available, the system can proceed to capture the drawing style of a particular artist. The proposed drawing style capturing algorithms are presented in Chapters, 6 and 7.
Chapter 6

Entire Face-Based Caricature Generation Approach

6.1 Introduction

The capabilities of artificial neural networks have been discussed in Chapter 4. This chapter proposes a novel example-based caricature generation system, which utilises the effective learning ability of neural networks to capture the drawing style of an artist. Experimental results and detailed analysis are provided to demonstrate that the proposed approach is capable of capturing the drawing style of an artist and is able to thus create photorealistic caricatures.

This chapter proposes a novel entire face-based approach to caricature generation. It is organised as follows:

6.2 An overview
6.3 Relationships among original image, caricature and mean face
6.4 Entire face-based caricature generation approach
6.5 Experimental results and analysis
6.6 Subjective test
6.7 Further experiments
6.8 Chapter summary and conclusions
6.2 An Overview

Caricature generation using artificial intelligence technologies has not been attempted in the past research (see Chapter 3). In this chapter, a novel caricature generation system that uses neural networks has been proposed and due to the approach adopted has been named as the "Entire Face-based Approach". This approach considers the entire human face as a single object and performs caricaturing of different facial components simultaneously, proving that neural networks are capable of capturing the drawing style of an artist. An overview of the whole approach can be presented as follows:

Firstly, a number of dataset preparation stages are required as described in Chapter 5, which include face selection, caricature drawing, normalisation, manual marking of feature points and mean face generation. Subsequently, within the training phase, the geometrical differences between an original facial image and the mean face (calculated based on the feature points) are fed to the neural network as inputs; while the geometrical differences between the corresponding caricature and the original image are considered as the outputs of the neural network. A feed-forward back-propagation neural network with one hidden layer has been used (see subsection 6.4.3). Once the training phase is completed, the geometrical differences between a validation image and the mean face are fed to the neural network for evaluation. Finally, the corresponding output, after converting to \((x,y)\) coordinates, is forwarded to a mesh warping module (see subsection 6.4.5) that deforms the original validation image into a caricature. A block diagram that summarises the proposed entire face-based automatic caricature generation algorithm is illustrated in figure 6-1. In the following sections, more detailed descriptions of the processes introduced above are provided.
Figure 6-1. The proposed entire face-based automatic caricature generation algorithm.
6.3 Relationships among Original Image, Corresponding Caricature and Mean Face

Once the mean face has been generated in the dataset preparation stages of Chapter 5, it will be manually marked with feature points. Corresponding feature points will also be marked manually on the original image and on the caricature (see subsection 5.5.4). Afterwards the deviations between the corresponding feature points of the original image, the caricature and the mean face can be estimated (see figure 3-24).

Even though the exact process of drawing a caricature from a facial image is hard to describe, previous research in psychology [25-26,31] have shown that it can be explained as follows:

Every caricaturist has a mean face in his/her mind, which is a result of the human psycho-visual system that works in a capacity similar to the ‘Morpher [103]’ explained in Chapter 5. This unconscious knowledge of the mean face gives the caricaturist the ability to identify distinctive features of a new facial image being viewed.

Let $\Delta S$ be the difference between an original face, $O$, and the mean face, $M$. Therefore,

$$\Delta S = O - M$$ (4)

Considering $\Delta S$, the artist then exaggerates the original image to form a caricature. The difference between the caricature, $C$, and its corresponding original image, $O$, is the change made by the artist in drawing the caricature, which is defined as $\Delta S'$. 

102
accordingly.

$$\Delta S' = C - O \quad (5)$$

In summary, when an artist sees the difference $\Delta S$, then he/she makes the change $\Delta S'$. Therefore the relationship between $\Delta S$ and $\Delta S'$ defines the artist's drawing rules that govern his/her drawing style. It is known that different artists have different styles of drawing caricature as the rules embedded in their subconscious minds are different (see section 2.3). Thus by capturing the relationship between $\Delta S$ and $\Delta S'$ from the drawings of an artist, the artist's drawing style can be summarised into a set of rules. This provides the ability to apply the above captured rules to a totally new image, generating a caricature with the artist's style embedded. However, the relationship between $\Delta S$ and $\Delta S'$ is always non-linear (see subsection 3.3.1 "Brennan's caricature generator") and the rules are difficult to describe in written language precisely (see subsection 3.3.3). As a result, an artificial intelligence based approach is adopted, i.e. the use of neural networks, to accomplish this task. Further explanations will be covered in section 6.4.
6.4 Entire Face-Based Caricature Generation Approach

6.4.1 Artificial Neural Networks

Artificial neural networks, a type of artificial intelligence technologies with proven learning ability has been presented and discussed in Chapter 4. Although alternative machine-learning techniques exist, e.g. Boolean expression learning, decision trees, statistical learning, etc [107], neural networks have been chosen as the underlying artificial intelligence technology due to the following reasons:

The main reason for utilising neural networks within the proposed research context is their effective ability of solving problems that are too complicated or fuzzy for conventional technologies, where an algorithmic solution is not yet available or too difficult to be found [82]. As an artificial neural network is an abstract of a human brain, it performs extraordinarily for problems that are good at been solved by a human, but not by a computer. Typical examples are data mining and pattern forecasting that involve recognition and analysis of trends from data provided, which are similar to the drawing style capturing task of the proposed system. Beside, a neural network is capable of learning from a training set by constructing an input-output mapping for the problem automatically. Therefore, an understanding of how the input is mapped to the output is not necessary, which is ideally suitable to be used for capturing the unexplainable relationship between $\Delta S$ and $\Delta S'$ (see section 6.3) in caricature generation. Moreover, a neural network has the ability to capture non-linear relationships from a training set. The non-linear equations embedded in neurons are able to compute and provide non-linear results (see subsection 4.4.4(vii)). This is suitable for capturing non-linear exaggerations of facial components in caricaturing as discussed in Chapter 3.
6.4.2 Preparation of Training and Validation Sets

The training set of a neural network consists of both input and output values. Therefore in the proposed approach, in order to capture the relationship between $\Delta S$ and $\Delta S'$, the input and output to the neural network should be $\Delta S$ and $\Delta S'$ respectively.

Figure 6-2 illustrates a simple example that can be used to establish a relationship between $\Delta S$ and $\Delta S'$. For the purpose of explanation, assume that each oval corresponds to a contour of an eye, and each is defined by eight feature points. Assume that the innermost oval is a mean eye, the one in the middle is an original eye and the outermost oval is a caricature eye. They are overlapped with each other by using the iris as the common reference point. In this example, the original eye is slightly bigger than the mean eye. Therefore the artist is expected to exaggerate the caricature eye in a non-linear way, during which both the size and the shape of the caricature eye can be changed.

![Overlapping Reference Point: Iris](image)

**Figure 6-2. Calculation of $\Delta S$ and $\Delta S'$.**
Chapter 6. Entire Face-Based Caricature Generation Approach

Specifically defining $\Delta S$ to be the separation between two corresponding feature points of the original image (say $O$) and the mean shape image (say $M$), their separation in $x$ and $y$ directions can be written as:

$$X_{\Delta S} = X_O - X_M$$  \hspace{1cm} (6) $$Y_{\Delta S} = Y_O - Y_M$$  \hspace{1cm} (7)

Similarly, defining $\Delta S'$ to be the separation between two corresponding feature points of the caricature (say $C$) and its corresponding original image, their separation in $x$ and $y$ directions can be written as:

$$X_{\Delta S'} = X_C - X_O$$  \hspace{1cm} (8) $$Y_{\Delta S'} = Y_C - Y_O$$  \hspace{1cm} (9)

After calculating the $\Delta S(x,y)$ and $\Delta S'(x,y)$ of all feature points, a table consisting of the training set entries is prepared. Table 6-1 illustrates a section of this table.

<table>
<thead>
<tr>
<th>FDP Number (see figure 5-6)</th>
<th>$X_{\Delta S}$ of the 1st Original image-Mean Pair</th>
<th>$X_{\Delta S'}$ of the 1st Original image-Caricature Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11.1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>11.2</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>11.3</td>
<td>-4</td>
<td>-1</td>
</tr>
<tr>
<td>4.4</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 6-1. An example of a part of a training set.
Collecting data for the project was an extremely difficult task as mentioned in Chapter 5. Based on the limited dataset obtained and used for experimentation, capturing the drawing style of an artist is a major challenge. Fortunately, the use of reliable neural networks, along with a cross-validation algorithm, provides a viable solution to the above limited dataset problem. Cross-validation is a common practice used for estimating generalisation error based on "re-sampling," where further samples are scarce or costly to obtain. The effectiveness of cross-validation was demonstrated to be superior for small datasets by Goutte in 1997 [108].

In the conventional neural network training approach, the dataset is divided into training, validation and test subsets, where the validation subset is used to choose the model parameters that achieve the highest generalisation level, whilst the test subset is used to measure the performance of the trained neural network with unseen data [77]. However, a cross-validation algorithm duplicates and partitions the dataset into a number of training subsets and validation subsets with various combinations, where cases that exist in a training subset are omitted in the corresponding validation subset. The same number of identical neural networks is then created, with each trained by a single instance of the training subsets and evaluated by the corresponding validation subset, in order to tackle the limited dataset problem. The advantage of using cross-validation algorithm is that it allows the entire dataset to participate in training. Consequently, the information embedded in the limited sample environment can be fully explored [108].

In this project, the cross-validation strategy is adopted to solve the problem of limited dataset. Two samples out of twelve, i.e. samples 11 and 12 of figure 5-15, were randomly selected as the first validation set, and the remaining samples, i.e. samples
1-10, are used as the training set. In the second cross-validation instance, two different samples were randomly chosen as the validation set, i.e. 2 and 7 of figure 5-15, whereas the remaining images (including 11 and 12) are used as the training set. The prepared training and validation sets are forwarded to the neural network training module to be presented in subsection 6.4.4.

6.4.3 Neural Network Architectures

Once the training set has been prepared, the next step is to define a neural network. In this project, the use of a feed-forward back-propagation network with only one hidden layer is proposed. The basic concepts of neural networks can be found in Chapter 4. The following discusses the reasons of choosing the proposed neural network type.

Neural networks can approximately be categorised into 18 types. However according to applications in which they are suitable to be used, they can be grouped into four major categories, namely, networks for classification and prediction, data association, data conceptualisation and data filtering [79]. Obviously, the network for classification and predication is the most appropriate category for the proposed research, as capturing the drawing style of an artist from his/her products is a pattern classification task and the subsequent drawing style imitation in the final output is, predication. As a result, the category of "networks for classification and prediction" was used for the proposed research.

The feed-forward back-propagation neural network is the most popular model in the classification and prediction category, due to its effectiveness, simplicity and fast excitability [79]. It is commonly known as the "universal function approximator" owing to its ability to teach itself, anything learnable [82]. Hence, a feed-forward
back-propagation neural network is adopted within the present context of this research. However, there is no standard rule to specify the number of hidden layers that a neural network is required to use. Minsky and Papert mathematically proved that a single layer ‘perceptron’ (without hidden layer) was insufficient to cope with classification tasks that were linearly inseparable, like the classes in XOR [109]. Therefore, multi-layer networks with one or two hidden layers appear in most practical applications. As the trial and error method is recognised as one of the best solutions to find out the number of layers that should be adopted, within the present context of this research a one hidden layer neural network is adopted for the purpose of simplicity.

The numbers of nodes used in the input, hidden and output layers are the same as the total number of FDPs proposed in the “Simple Face Model” of Chapter 5, i.e. 46 (see table 5-1), since the $\Delta S$ of each FDP enters the neural network as an independent input whilst the corresponding $\Delta S'$ forms the output. The number of nodes in the hidden layer is designed according to “the square root of the product of the number of input nodes multiplied by the number of output nodes” as proposed by Masters [79] and has been discussed in Chapter 4. The results of neural network using various numbers of hidden neurons will also be discussed in section 6.7. Table 6-2 summarises the details of neural network architecture and training parameters used.
### Architecture and Parameters

<table>
<thead>
<tr>
<th>的选择项</th>
<th>选择</th>
<th>说明</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Neural Networks Required per Artist</td>
<td>2 (x,y coordinates are trained separately)</td>
<td>2个神经网络用于每个艺术家的x、y坐标独立训练</td>
</tr>
<tr>
<td>Neural Network Type</td>
<td>Feed-forward Back-propagation</td>
<td>前向传播反向传播神经网络</td>
</tr>
<tr>
<td>Training Function</td>
<td>Levenberg-Marquardt</td>
<td>李文博-马夸特训练函数</td>
</tr>
<tr>
<td>Performance Validation Function</td>
<td>Mean Squared Error (MSE)</td>
<td>均方误差性能验证函数</td>
</tr>
<tr>
<td>Performance Goal</td>
<td>0 (default)</td>
<td>性能目标：0（默认）</td>
</tr>
<tr>
<td>Minimum Gradient</td>
<td>1e-010 (default)</td>
<td>最小梯度：1e-010（默认）</td>
</tr>
<tr>
<td>Maximum Number of Epochs</td>
<td>100000</td>
<td>最大迭代次数：100000</td>
</tr>
<tr>
<td>Maximum Fail</td>
<td>5 (default)</td>
<td>最大失败次数：5（默认）</td>
</tr>
<tr>
<td>Memory Reduction</td>
<td>10</td>
<td>内存减少：10</td>
</tr>
<tr>
<td>Mu Initial $\mu$</td>
<td>0.001 (default)</td>
<td>$\mu$初值：0.001（默认）</td>
</tr>
<tr>
<td>Mu Decrease</td>
<td>0.1 (default)</td>
<td>$\mu$递减：0.1（默认）</td>
</tr>
<tr>
<td>Mu Increase</td>
<td>10 (default)</td>
<td>$\mu$递增：10（默认）</td>
</tr>
<tr>
<td>Mu Maximum</td>
<td>10000000000 (default)</td>
<td>$\mu$最大值：10000000000（默认）</td>
</tr>
<tr>
<td>Show</td>
<td>25 (default)</td>
<td>显示：25（默认）</td>
</tr>
<tr>
<td>Time</td>
<td>Infinite (default)</td>
<td>时间：无限（默认）</td>
</tr>
<tr>
<td>Momentum</td>
<td>No (default)</td>
<td>动量：否（默认）</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.01 (default)</td>
<td>学习率：0.01（默认）</td>
</tr>
<tr>
<td>Number of Layers</td>
<td>3</td>
<td>层数：3</td>
</tr>
<tr>
<td>Hidden Layer Transfer Function</td>
<td>Logistic Sigmoid (logsig)</td>
<td>隐藏层传输函数：逻辑 sigmoid</td>
</tr>
<tr>
<td>Output Layer Transfer Function</td>
<td>Pure Linear (purelin)</td>
<td>输出层传输函数：纯线性 (purelin)</td>
</tr>
<tr>
<td>Number of nodes in the input layer</td>
<td>46</td>
<td>输入层节点数：46</td>
</tr>
<tr>
<td>Number of nodes in the hidden layer</td>
<td>46</td>
<td>隐藏层节点数：46</td>
</tr>
<tr>
<td>Number of nodes in the output layer</td>
<td>46</td>
<td>输出层节点数：46</td>
</tr>
</tbody>
</table>

|Table 6-2. Neural Network Architecture and Training Parameters.|
6.4.4 Neural Network Training and Validations

After the neural network is constructed, the training process can be commenced. Several experiments were carried out to decide upon the most suitable training function to be used. MATLAB and its ANN toolbox were used as the programming language/environment [112]. The Levenberg-Marquardt [110] algorithm without momentum was found to be the best approach. It is a mathematical procedure used to find the minimum of a function that is a sum of squares of nonlinear functions. Subsequently the weights and biases of the network are backward adjusted automatically. This algorithm is likely the fastest method for training moderate-sized feed-forward neural networks that are up to several hundred weights [110]. The mean squared error was used as the performance validation function and the performance goal was set to zero. The Mean Squared Error (MSE) is defined as:

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (T - \theta)^2
\]  

(10)

where \( n \) = sample size, \( T = \Delta S' \) from training set, \( \theta = \text{modelled } \Delta S' \).

The training stopped once the performance was minimised to the goal or the gradient of performance was less than the minimum gradient parameter, which means the error slope is close to zero and further training is unlikely to reduce the error by any significant amount [111]. Note that in the neural network toolbox of MATLAB, the end of training depends on the runtime mean squared error. This means that the number of epochs (cycles) to be trained cannot be specified by the users.

Putting a neural network into practice immediately after training is unwise as its accuracy and capabilities should first be evaluated and scrutinised, this process is known as validation. The \( \Delta S \) describing (at each FDP) the differences of a new
original facial image (from validation subset) from the mean face is fed as input to the
trained neural network for validation. The neural network will then generate $\Delta S'$ that
describes (at each FDP) the differences of predicted caricature image from the
original validation image. As the FDPs of the original validation image are known, the
above differences can be used to obtain the XY coordinates that define the
corresponding caricature image, which are then forwarded to the final stage, i.e. the
mesh warping. The experimental validation results will be fully discussed in section
6.5. Further, results of neural networks with different architectures and parameters
will also be illustrated in section 6.7 for comparison purposes.

6.4.5 Mesh Warping
Once the (x,y) coordinate pairs defining the FDPs of both original image and the
caricature to be generated are ready, they will enter a mesh warping module that
converts the original image into its corresponding caricature. The idea of a mesh
warping algorithm [64] (see Chapter 4) is to deform one image (i.e. original) towards
another (i.e. caricature). The mesh warping algorithm of [64] was specifically used for
all caricature generation experiments of the proposed research. Note that similar
algorithms are at present a common inclusion in most picture editing software
packages [87-88].

Within the proposed approach, the (x,y) coordinates of the FDPs of the original image
form the source face mesh and the XY coordinates of feature points generated from
the neural network form the target face mesh. The warping module uses the source
mesh and warps the original image towards the target mesh. Hence a caricature of the
original face with the drawing style of the artist (due to being warped towards the
target mesh above) can be generated.
6.5 Experimental Results and Analysis

Several experiments were carefully designed and carried out to demonstrate the ability of the proposed approach to learn the drawing style of a caricaturist and automatically produce photorealistic caricatures that are embedded with the unique drawing style of the artist. Three professional caricaturists were invited to draw caricatures of twelve male facial images from the AR face database (see subsection 5.5.2 for details and figure 5-15 for the results). In order to further reduce the need for excessive computer resources during the neural network training process, X and Y coordinates were trained separately in two identical neural networks, which were constructed based on the parameters provided in table 6-2.

Subsequent to the training of the neural network described in subsection 6.4.4, the $\Delta S'(x,y)$ of FDPs of validation images, obtained from the trained neural networks’ outputs are converted to (x,y) coordinates and forwarded to the mesh warping module (see subsection 6.4.5) for the creation of the photorealistic caricatures, which are then compared with caricaturist’s drawings for validation.

It can be shown that analysing the captured drawing style is a nonfigurative task, as a careful consideration concludes that visual comparisons of the generated caricatures would provide a more direct and artistic evaluation of the proposed system, instead of the common statistical neural network performance analysis methods, such as the Mean Squared Error (MSE), the Coefficient of Efficiency (CE) and the Standard Error of the Estimate (SE) [77]. As a result, a visual analysis of the first cross-validation experiment (i.e. images 11 and 12) by the author will be fully discussed in section 6.5, which is followed by a statistical analysis of the results of a subjective test in section 6.6. The experimental results of the second cross-validation instance (i.e. images 2
and 7) will be presented in section 6.7.

As the proposed research focuses on the shape exaggeration style of an artist instead of the texture (see subsection 3.3.3) changes used, it can be reasonably assumed that the freehand and photorealistic caricatures are comparable with each other. The experimental results for the three artists named artist-1, artist-2 and artist-3 are separately presented in subsections 6.5.1, 6.5.2 and 6.5.3 respectively. Note that (i) and (ii) separate the results obtained from validation images 11 and 12 respectively.
6.5.1(i) The First Validation Experiment of Artist-1:

Figure 6-3(i). (a) Original image 11 (validation case 1). (b) Caricature drawn by Artist-1. (c) Caricature generated by the proposed entire face-based approach.

Figure 6-3(i) demonstrates the result of the first validation experiment of artist-1, by comparing the caricature drawn by the first artist, (b), with the caricature generated by the proposed system, (c), it can be shown that some of the drawing styles have been captured successfully.

First of all, the height of the forehead in (b) is shortened when compared with (a). A very similar forehead distortion is illustrated in (c). Besides, similar exaggerations of the nose appear in both (b) and (c). Not only are the width of nose increased, but also the length are elongated. Moreover, the changes of the mouth in (b) and (c) are very close to each other. Both of them are stretched in width but not in height.

Finally, the shapes of the eyes and eyebrows in (c) are only slightly changed when compared with (a). These are similar to (b) as artist-1 did not modify them vigorously. Thus it can be argued that the proposed caricature generation approach has successfully picked up a particular trait of the artist's drawing style.
However, the shapes of the ears and the face in (c) do not have the same exaggeration style as in (b), these could due to insufficient feature points were used in training, and therefore these two components cannot be captured and represented accurately. Note that both hair and neck are not considered in the current system as they are not involved in the proposed "Simple Face Model". Therefore changes to the hair and neck cannot be captured by the present design of the proposed automatic caricature generation system.

6.5.1(ii) The Second Validation Experiment of Artist-1:

Figure 6-3(ii). (a) Original image 12 (validation case 2). (b) Caricature drawn by Artist-1. (c) Caricature generated by the proposed entire face-based approach.

Figure 6-3(ii) illustrates the results of the second validation experiment in which the caricature drawn by the first artist, i.e. (b), is compared with the caricature generated by the proposed approach, i.e. (c). It can be argued that satisfactory results are demonstrated.

Firstly, the ways exaggerating the nose in (b) and (c) are very close to each other, specifically both of them are slightly bigger than the original image (a). On top of this, both eyebrows and eyes of (c) are slightly wider than (a); these changes also match
Chapter 6. Entire Face-Based Caricature Generation Approach

the drawing of artist-1 in (b). Furthermore, the size of mouth of figures (b) and (c) are both wider than of (a), as the shapes remain unchanged. Moreover, the success of drawing style capturing is very obvious at the chin of figures (b) and (c), where it is elongated heavily.

Finally, similar exaggeration styles of ears appear in both (b) and (c), where shapes and exaggeration ratios are almost the same. These illustrates that non-linear exaggerations can be imitated by the proposed system. Note that the exaggeration ratio of right ear is larger than of the left ear, as the original image, (a), shows an unbalanced pair of ears when the subject is facing the camera.

Unfortunately, the height of the forehead in (b) cannot be imitated by the generated caricature (c). This can be explained by the proposed system incorrectly picked up the forehead drawing style of the 10th original image by artist-1, as illustrated in column 2 of figure 5-15.

6.5.2(i) The First Validation Experiment of Artist-2:

![Figure 6-4(i). (a) Original image 11 (validation case 1). (b) Caricature drawn by Artist-2. (c) Caricature generated by the proposed entire face-based approach.](image)
In order to further demonstrate the validity of the proposed drawing style capture algorithm, experiments concerning to artist-2's caricaturing products were carried out. A methodology identical to that adopted in section 6.4 for the first artist, was applied on the drawings of the second artist.

Figure 6-4(i) presents the result of the first validation experiment of artist-2. It illustrates that the proposed automatic caricaturing system has been successful in capturing the drawing style of the second artist as well. The heavily elongated face in (c) is similar to the change made by artist-2 in (b), though the degree of exaggeration is more than what is expected. This can be explained by the specific trait of the second artist, which appears to be elongating faces in the vertical direction in caricaturing than in the cases of other two artists. Note that the identification of the above trait requires careful comparison of the above artist's caricatures to the caricatures of the two other artists illustrated in figure 5-15.

Besides, the proposed approach precisely predicted the exaggeration of the nose by the second artist in (b) and reveals its successful capture in (c). Finally, the slightly exaggerated mouth and elongated ears in (c) also match the drawings of (b).
6.5.2(ii) The Second Validation Experiment of Artist-2:

Figure 6-4(ii). (a) Original image 12 (validation case 2). (b) Caricature drawn by Artist-2. (c) Caricature generated by the proposed entire face-based approach.

Figure 6-4(ii) illustrates the result of the second validation experiment of artist-2. It is seen that in the caricature produced by the proposed system, i.e. (c), the generated face is elongated when compared with the original, i.e. (a). A similar facial shape exaggeration is illustrated in the caricature (b) that is drawn by the second artist. However, the extent of elongation in (c) is more than that illustrated in (b), once again sharing a drawing style similar to that explained in subsection 6.5.2(i).

Added to this, both size and shape changes of the nose in (c) are very close to that of (b), proving the success of the drawing style capture of the nose. Furthermore, the slightly caricatured eyebrows, the widened eyes and the unchanged mouth illustrate similar variations in (b) and (c), as compared to those in (a).

Although a slightly tilted face and a dent at the chin cannot be seen in (c), a careful investigation revealed that these two drawing styles do not appear in the training set (see column 3 of figure 5-15). As a result, the proposed system has failed to imitate them, proving the need of a larger training dataset per given artist.
6.5.3(i) The First Validation Experiment of Artist-3:

Figure 6-5(i). (a) Original image 11 (validation case 1). (b) Caricature drawn by Artist-3. (c) Caricature generated by the proposed entire face-based approach.

The methodology adopted in subsections 6.5.1 and 6.5.2 to analyse the caricatures of artists 1 and 2 is applied in this section to analyse the caricatures drawn by the third artist, with the aim of further proving that the proposed algorithm is able to successfully capture the drawing style of an artist.

Figure 6-5(i) demonstrates the result of the first validation experiment of artist-3. Once again, the caricature generated by the proposed system, i.e. (c), illustrates similar traits to that of the caricature drawn by the artist, i.e. (b). First of all, when comparing the face and ears in (b) and (c), the elongations of these components show that the drawing style of the third artist has been captured. However, the overall shapes of these two components are still not able to exactly imitate those drawn by artist-3 in (b); these provide a clue that more feature points are required to represent these components. Moreover, the eyes and nose in (c) are slightly shrunk and compare well with the modifications made by the artist to these components as illustrated in (b). Further, the exaggerated eyebrows and the slightly widened mouth in (c) are similar to the corresponding components of the artist’s caricaturing product of (b).
6.5.3(ii) The Second Validation Experiment of Artist-3:

![Figure 6-5(ii)](image)

Figure 6-5(ii). (a) Original image 12 (validation case 2). (b) Caricature drawn by Artist-3. (c) Caricature generated by the proposed entire face-based approach.

Figure 6-5(ii) illustrates the result of the second validation experiment related to the third artist, further proving the operational success of the proposed automatic caricature generation system.

First of all, when comparing the caricature generated by the proposed system, i.e. (c), and the caricature drawn by the artist, i.e. (b), the extent of the facial shape elongation of (c) evidently demonstrates the prediction ability of the proposed drawing style capture approach. Note that the exaggeration of face is specifically emphasised on the area between the mouth and the chin. However, the shape of the lower jaw illustrated in (c) did not quite pick up the way that the jaw was caricatured by the artist in (b). The reason for this is the insufficient amount of FDPs used to represent the area closer to the lower jaw in the proposed “Simple Face Model” (see Chapter 5). An enhanced approach to automatic caricature generation with increased number of FDPs in the jaw area will be covered in Chapter 7.

Besides, the elongations of ears in the automatically generated caricature of (c) are
similar to that made in the caricature of the second artist illustrated in (b). It is further noted that the other facial components in (c), such as eyebrows, eyes, nose and mouth, are slightly modified or remain unchanged as demonstrated by (b), which can also be considered as a part of capturing the drawing style of the third artist.
6.5.4 Further Analysis of Results

In the analysis carried out above, the automatically generated caricatures by the proposed system were directly compared with the corresponding artist’s caricatures. More convincing visual proof of the drawing style capturing ability of proposed automatic caricature generation system can be provided by comparing the automatically generated caricatures of the neural network trained on the caricaturing products of a single artist, with the caricatures drawn by all three artists for the same original facial image.

*Result comparisons of the system trained with Artist-1’s drawings:*

<table>
<thead>
<tr>
<th>Case</th>
<th>Original Image</th>
<th>Artist-1 Drawing</th>
<th>Artist-2 Drawing</th>
<th>Artist-3 Drawing</th>
<th>Generated by Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
<td>(d)</td>
<td>(e)</td>
</tr>
<tr>
<td>2</td>
<td>(f)</td>
<td>(g)</td>
<td>(h)</td>
<td>(i)</td>
<td>(j)</td>
</tr>
</tbody>
</table>

Figure 6-6. Validation results of the proposed system when trained on the caricatures of Artist-1. Row 1 and 2 are the first and the second validation cases respectively.
Chapter 6. Entire Face-Based Caricature Generation Approach

Figure 6-6 illustrates the results for the two validation images (a) and (f), when the neural network was trained with the caricaturing products of only the first artist. Columns 2, 3 and 4 of the figure illustrate the caricatures drawn by the three artists for the two validation images, whereas the column 5, i.e. images (e) and (j), illustrates the caricatures produced by the proposed system. In the case of the first validation image, i.e. (a), when comparing the computer generated caricature (e), with caricatures drawn by the three artists, it looks remarkably closer to artist-I’s drawing i.e. (b), than to the others, i.e. (c) and (d). A careful comparison of columns 1 (original images) and 2 (caricatures drawn by artist-I) of figure 5-15 reveals that the horizontal direction of the forehead area is exaggerated less as compared to the area near the cheek bone. Thus forehead area appears narrower as compared to cheek bone area in general, which can be identified as a trait of artist-I. This style has been preserved in the validation results of (e) and (j).

Similarly, for the second validation image, i.e. (f), the computer generated caricature, (j), is more similar to artist-I’s drawing (g), than the others, i.e. (h) and (i), as the traits of forehead and cheeks mentioned above have been maintained. Although (g) and (i) share some similarities, careful observations reveal that the shape of the jaw in (j) is more similar to that of (g) than to that of (i). Besides, the manner in which the ears are exaggerated in this validation image further supports the fact that the automatically generated caricatures resemble the drawing style of artist-I.

In general it is observed that the unique facial features tend to be significantly exaggerated by artist-I as compared to what have been done by the other two artists (see figure 5-15). As a result, capturing the drawing style of artist-I is expected to be the easiest amongst the three artists.
### Result comparisons of the system trained with Artist-2's drawings:

<table>
<thead>
<tr>
<th>Case</th>
<th>Original Image</th>
<th>Artist-1 Drawing</th>
<th>Artist-2 Drawing</th>
<th>Artist-3 Drawing</th>
<th>Generated by Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
<td>(d)</td>
<td>(e)</td>
</tr>
<tr>
<td>2</td>
<td>(f)</td>
<td>(g)</td>
<td>(h)</td>
<td>(i)</td>
<td>(j)</td>
</tr>
</tbody>
</table>

Figure 6-7. Validation results of the proposed system when trained on the caricatures of Artist-2. Row 1 and 2 are the first and second validation cases respectively.

Figure 6-7 illustrates the validation results of the proposed system when the neural network was trained only on the caricaturing products of artist-2. It reveals that the computer generated caricatures look more similar to the drawings of the second artist than to the drawings of the others, in both validation cases. In the first validation case, the computer generated caricature (e) looks closer to the caricature of artist-2, i.e. (c) than to (b) or (d), as it's elongated face with an exaggerated nose appears to dominate the drawing style of artist-2. In the second validation experiment, the elongated nose and the shrunk mouth in the computer generated caricature (j) clearly demonstrate that the drawing style has been captured from that of the second artist, as illustrated by (h), but not by the others as illustrated by (g) or (i).
Chapter 6. Entire Face-Based Caricature Generation Approach

**Result comparisons of the system trained with Artist-3’s drawings:**

<table>
<thead>
<tr>
<th>Case</th>
<th>Original Image</th>
<th>Artist-1 Drawing</th>
<th>Artist-2 Drawing</th>
<th>Artist-3 Drawing</th>
<th>Generated by Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="a" alt="Image" /></td>
<td><img src="b" alt="Image" /></td>
<td><img src="c" alt="Image" /></td>
<td><img src="d" alt="Image" /></td>
<td><img src="e" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="f" alt="Image" /></td>
<td><img src="g" alt="Image" /></td>
<td><img src="h" alt="Image" /></td>
<td><img src="i" alt="Image" /></td>
<td><img src="j" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 6-8. Validation results of the proposed system when trained on the caricatures of Artist-3. Row 1 and 2 are the first and second validation cases respectively.

Finally, in the validation results of artist-3, both experiments show that the proposed system generated caricatures with a drawing style that is closest to artist-3’s caricatures than to that of artist-1 or 2. In the first validation case, the similarity between the computer generated caricature, (e), and the artist-3’s caricature, (d), is the best. It is further observed that the artistic trait of an elongated face with a shrunk nose that exists in the computer generated caricature, (e), is only apparent in (d) but not in (b) or (c). Similar results are illustrated in the second validation case (row 2 of figure 6-8), where the drawing style of the computer generated caricature, (j), apparently is closest to (i) as the modifications made to the eyes, nose, chin, mouth and ears are very similar. Further though it appears as the hand-drawn caricatures (g) and (i) are similar to each other, a careful observation reveals that the drawing style
embedded in (j) best fits to that of (i) as the shapes of their faces are almost the same.
All above evidence reveal that the proposed approach is able to reproduce the drawing
traits of artist-3 in the automatically generated caricature outputs after being trained
by his caricaturing products.

It was stated that in general the artist-1's drawing style appears to have the most
significant exaggerations and artist-3's, the least. A careful comparison of figures 6-6,
6-7 and 6-8 reveal that computer generated images of artist-1's caricatures have the
highest similarity to the hand-drawn caricatures, while artist-3's have the least.

Even though the proposed system cannot completely predict and generate caricatures
that are exactly the same as the drawings of the artists, the above comparisons
demonstrate that a certain extent of the drawing rules has been successfully captured
after trained with the caricatures of a given artist. In conclusion it can be stated that all
the validation experiments support fact that the proposed caricaturing system is able
to satisfactorily capture and reproduce the drawing style of a given artist.
6.6 Subjective Test

The experimental results analysed above was a thorough visual analysis of the caricatures by the author who has significant understanding of the caricaturing process, and the strengths and weaknesses of the existing and proposed approaches to automatic caricature generation. In order to further support the claim of proposed system’s ability to successfully capture the drawing style of an artist, a scientific subjective validation test is required to evaluate the proposed system, without the direct influence of the author’s judgments. However, identifying the style of a piece of art has not been a well researched topic in past research. Only few efforts have been made on identifying the style of music, poetry and handwriting [113-115] of a given individual, whereas no research efforts have investigated the identification of the caricatures of an artist so far. For the conventional drawings (non-caricature), most of the attempts focused on texture classification and reproduction, e.g. oil painting, pencil sketch, etc [116-118], instead of the drawing style of a particular artist. David Stork [120-121] has done some work on judging whether early Renaissance paintings were originals or fakes. However, the idea proposed in [120-121] relies on optical equipment for analysis, which is far from the methodology adopted by this research.

In view of the above factors, a subjective test for drawing style validations was carried out. The experimental results from section 6.5 were presented to 46 volunteers, who are novice to caricaturing, but would be able to comment on the quality of the output results based on a general visual analysis.

A questionnaire was designed to investigate the ability of the proposed approach in general to capture the drawing style of an artist and hence automatically reproduce caricatures with the same style. All subjects were shown pictures of two original faces
(i.e. images 11 and 12 of figure 5-15), hand-drawn caricatures of the faces by the three artists (with ownership assigned) and the computer generated caricatures when the entire face-based approach was used, without the ownership being assigned. The task of the subjects was to select the owner of each unclassified computer generated caricature, where the images from two validation cases were tested separately. Note that the aim of the subjective test was to identify the computer generated caricatures from the provided artists’ products based on the similarity of the drawing styles, instead of finding the most impressive caricature amongst the caricatures produced by the three artists. Figure 6-9 briefly illustrates the images used in the questionnaire and its basic nature. A complete questionnaire can be found in Appendix B.

**Question 1:**

<table>
<thead>
<tr>
<th>Case</th>
<th>Original image</th>
<th>Artist-1 drawing</th>
<th>Artist-2 drawing</th>
<th>Artist-3 drawing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image" alt="Original Image" /></td>
<td><img src="image" alt="Artist-1 Drawing" /></td>
<td><img src="image" alt="Artist-2 Drawing" /></td>
<td><img src="image" alt="Artist-3 Drawing" /></td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
<td></td>
</tr>
</tbody>
</table>

**Computer Generated Caricatures (in random order)**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Computer Generated Caricature" /></td>
<td><img src="image" alt="Computer Generated Caricature" /></td>
<td><img src="image" alt="Computer Generated Caricature" /></td>
</tr>
</tbody>
</table>

**Answer:**

Figure 6-9 (i). Subjective test of the first validation case. Participants were invited to match each artist's drawing, i.e. caricatures of (a), (b) and (c) of row 1 above, to a computer generated image in row 2.
Question 2:

<table>
<thead>
<tr>
<th>Case</th>
<th>Original image</th>
<th>Artist-1 drawing</th>
<th>Artist-2 drawing</th>
<th>Artist-3 drawing</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><img src="image" alt="Original image" /></td>
<td><img src="image" alt="Artist-1 drawing" /></td>
<td><img src="image" alt="Artist-2 drawing" /></td>
<td><img src="image" alt="Artist-3 drawing" /></td>
</tr>
</tbody>
</table>

Answer:

Figure 6-9 (ii). Subjective test of the second validation case.

The answers to the above questionnaire are summarised and graphically illustrated in figure 6-10, where case-1 and case-2 represent the two validations cases, i.e. results obtained when testing caricatures 11 and 12 respectively. The results illustrate that more than 75% of the subjects were able to correctly determine the ownership of each computer generated caricature when using the entire face-based approach. Due to the unique vertically elongated drawing style adopted by artist-2 that was identified and discussed in subsection 6.5.2, his computer generated caricatures have been identified more accurately, on the average, with artist-1’s caricatures coming a close second (due to acute feature exaggerations used) and artist-3’s, last (due to minimal exaggerations used in his drawing style).
Chapter 6. Entire Face-Based Caricature Generation Approach

The subjective test results obtained above were further analysed using statistical decision theory [119], which helps making decisions about populations on the basis of sample information. In this extended analysis, the decisions of “whether the obtained results of subjective test (presented above) are due to chance (completely guessing) or are based on the proper identification of similarity of drawing styles” were further investigated.

Prior to reaching decisions, assumptions which may or may not be true, about the probability distributions of the populations that are formally known as “statistical hypotheses” are required. If a hypothesis is rejected when it should be accepted statistical decision theory states that a so-called Type I error has been made. On the other hand, if a hypothesis is accepted when it should be rejected, it is said that a so-called Type II error has been made. Therefore a wrong decision or error in judgment occurs in either of the above two cases [119]. Only the hypotheses in which both expectation and statistical decision results match are considered as correct.
In testing a given hypothesis, the maximum probability of Type I error that is accepted to be risked is called the *level of significance* of the test, which is usually denoted by $\alpha$. Levels of significance of 0.05 and 0.01 are common practices and define the percentages of confidence that the right decisions have been made. For example, if the level of significance of 0.05 is chosen in testing a hypothesis, it means that 5 chances out of 100 would reject the hypothesis when it should be accepted. In other words, a 95% confidence level of the right decision made is obtained. For the level of significance 0.01, a 99% confidence is held.

In order to present the idea above, a normal distribution is commonly used to illustrate the sampling distribution of a statistic $S$. When the mean and the standard deviation are denoted by $\mu_s$ and $\sigma_s$ respectively, the distribution of the standardised variable (z score) is the standardised normal distribution [119], and is given by,

$$z = (S - \mu_s) / \sigma_s$$  \hspace{1cm} (11)

![Figure 6-11. Standardised normal distributions. (a) one-tailed test. (b) two-tailed test.](image-url)
Figure 6-11 shows the standardised normal distribution graphs of one-tailed and two-tailed tests in (a) and (b) respectively. A one-tailed test concentrates on an increase or decrease in the parameter whereas a two-tailed test considers any change in the parameter (which can be either increased or decreased) as a range is provided.

If a hypothesis is true with a 0.05 level of significance, then the z score of a random sample of statistic $S$, will have a 95% confidence and will lie within the area of the normal curve excluded by the critical regions, i.e. the range from -1.96 to 1.96 for the two-tailed test (see figure 6-11b) and $>-1.645$ or $<1.645$ for the one-tailed test (see figure 6-11a). However, if any sample with z score lies outside the range, i.e. within the critical regions, a conclusion can be made that such an event could happen with probability of only 0.05 provided that the given hypothesis was true [119].

Table 6-3 provides the critical values of z for both one-tailed and two-tailed tests at different levels of significance, and serves as a reference to the subsequent tests.

<table>
<thead>
<tr>
<th>Level of Significance $\alpha$</th>
<th>.10</th>
<th>.05</th>
<th>.01</th>
<th>.005</th>
<th>.002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical Values of z for One-Tailed Tests</td>
<td>-1.28 or</td>
<td>-1.645 or</td>
<td>-2.33 or</td>
<td>-2.58 or</td>
<td>-2.88 or</td>
</tr>
<tr>
<td>Critical Values of z for Two-Tailed Tests</td>
<td>1.28</td>
<td>1.645</td>
<td>2.33</td>
<td>2.58</td>
<td>2.88</td>
</tr>
<tr>
<td>Critical Values of z for One-Tailed Tests</td>
<td>-1.645 and</td>
<td>-1.96 and</td>
<td>-2.58 and</td>
<td>-2.81 and</td>
<td>-3.08 and</td>
</tr>
<tr>
<td>Critical Values of z for Two-Tailed Tests</td>
<td>1.645</td>
<td>1.96</td>
<td>2.58</td>
<td>2.81</td>
<td>3.08</td>
</tr>
</tbody>
</table>

Table 6-3. Critical values of z for one and two-tailed tests.

The results of the subjective test presented above are analysed using statistical decision theory and is detailed as follows:
If \( p \) is the probability of participants matching the caricature drawn by an artist with the corresponding computer generated caricature correctly, one of the following two hypotheses has to be accepted:

\[ H_0: p = .33, \] and participants randomly match a computer generated caricature to one of the three hand-drawn caricatures provided, i.e. results are due to chance.

\[ H_1: p > .33, \] and participants match the caricature drawn by an artist with the corresponding computer generated caricature based on similarity of style, i.e. the drawing style of the artist is embedded in the generated caricature.

If the hypothesis \( H_0 \) is true, the mean, \( \mu \), and the standard deviation, \( \sigma \), of the number of caricature pairs matched correctly is given by

\[
\mu = Np \\
\sigma = \sqrt{Npq}
\]

\[ (12) \]

\[ (13) \]

where \( N \) is the sample size, \( p \) is the population proportion of successes and \( q \) is equal to \( 1-p \). Therefore:

\[
\mu = 46(0.33) = 15.33 \\
\sigma = \sqrt{46(0.33)(0.67)} = 3.20
\]

Due to the fact that incorrect matching of caricature pairs is not of interest but rather only correct matching, the one-tailed test is selected to examine the formulated hypothesis. A significance level of 0.01 is also adopted as it provides a considerably
precise assessment to the given test. For a one-tailed test at a level of significance of 0.01, the critical value of $z$ is 2.33 (see table 6-3). Thus the decision rule or test of significance is:

1. If the $z$ score observed is greater than 2.33, the results are significant at the 0.01 level and participants match the caricature drawn by an artist with the corresponding computer generated caricature based on similarity of the drawing styles.

2. If the $z$ score is less than 2.33 the results are due to chance, i.e. not significant at 0.01 level.

By substituting the number of participants that were able to make a correct match $S$, (for each caricature) into equation (11), their $z$ scores can be calculated and can be tabulated as in table 6-4.

<table>
<thead>
<tr>
<th>Validation</th>
<th>Artist-1</th>
<th>Artist-2</th>
<th>Artist-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>Case 2</td>
<td>Case 1</td>
<td>Case 2</td>
</tr>
<tr>
<td># of participants matched the correct answer</td>
<td>39</td>
<td>35</td>
<td>38</td>
</tr>
<tr>
<td>$&gt; 2.33$</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 6-4. $z$ score calculations of the participants who were able to make a match.

As all the $z$ scores are greater than 2.33, decision (1) holds in all cases, i.e. it can be concluded at 0.01 level that participants accurately matched the caricatures drawn by the artists with the computer generated caricatures, based on similarity of drawing
styles. As a result, the hypothesis that “subjects have been able to correctly match the artist’s caricature with that of the computer generated caricature using the proposed algorithm”, is found to hold at a confidence level of over 99%, for all caricatures drawn by all three artists.
6.7 Further Experiments

Additional experiments were designed to evaluate the performance of neural networks, when trained with different parameter selections.

6.7.1 Different Number of Neurons in the Hidden Layer

The trial and error method is widely recognised as one of the most feasible solutions for finding out the most effective number of hidden neurons (see subsection 4.4.4) that should be used within a layer, by a neural network. Therefore the use of different numbers of neurons, in a single hidden layer, has been experimented within the research context of the project. Figure 6-12 illustrates the experimental results of the first cross-validation experiment of artist-1. Note that columns 2, 3, 4 and 5 represent the results of neural networks that use 10, 23, 34 and 46 hidden neurons, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>10 neurons</th>
<th>23 neurons</th>
<th>34 neurons</th>
<th>46 neurons</th>
<th>Caricature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Original" /></td>
<td><img src="image2" alt="10 neurons" /></td>
<td><img src="image3" alt="23 neurons" /></td>
<td><img src="image4" alt="34 neurons" /></td>
<td><img src="image5" alt="46 neurons" /></td>
<td><img src="image6" alt="Caricature" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image7" alt="Original" /></td>
<td><img src="image8" alt="10 neurons" /></td>
<td><img src="image9" alt="23 neurons" /></td>
<td><img src="image10" alt="34 neurons" /></td>
<td><img src="image11" alt="46 neurons" /></td>
<td><img src="image12" alt="Caricature" /></td>
</tr>
</tbody>
</table>

Figure 6-12. Experimental results of using different numbers of hidden neurons.
generated caricatures as the number of neurons in the hidden layer increases, when comparing caricatures of columns 2, 3, 4 and 5 with the caricature drawn by the first artist, i.e. the caricature of column 6. For both validation cases, the best quality caricature is the one illustrated in column 5, i.e. when using 46 neurons. This demonstrates that the observation, "the required number of artificial neurons in the hidden layer is normally near the square root of the product of the number of input nodes multiplied by the number of output nodes" of Masters in [79] (see subsection 4.4.4) applies well to the proposed system. This observation was used in deciding the number of neurons in the single hidden layer of the neural network used within the main experiments of this Thesis presented in sections 6.5 and 6.6.

6.7.2 Different Neural Network Training Algorithms and Parameters

In the proposed system, the training function Levenberg-Marquardt was adopted (see table 6-2). However, the application of this function in MATLAB does not accept a momentum parameter, which is a common approach to overcome obstacles such as local minima in the error surface during training [77]. Therefore the use of several training algorithms with and without momentum, such as Gradient Descent, Random Order Incremental Training and One Step Secant Back-Propagation [112], were investigated. Unfortunately, none of these functions was able to finish the training with performance minimised to the goal or the gradient of performance less than the minimum gradient parameter (see subsection 6.4.4). In other words, all above training functions failed in training as their mean squared errors obtained were far from zero. Consequently, the training routines were stalled when the maximum number of epochs (training cycles), i.e. 100000 (see table 6-2), was reached. As a result, the use of the above alternative training functions within the neural network adopted was deemed not be eligible for consideration in the subsequent validation stage.
6.7.3 The Experimental Results of the Second Cross-Validation Instances

In the second instance of cross-validation experiments, the same methodology as adopted in the first instance was used (see section 6.4) except that the original images 2 and 7 of figure 5-15 were used for validations, whereas images 11 and 12 of figure 5-15 were included within the training set. The results for all three artists are illustrated in figures 6-13, (i) – (iii). Note that cases 1 and 2 represent validation experiments of original images 2 and 7, respectively.

The Second Validation Instance of Artist-1:

<table>
<thead>
<tr>
<th>Case</th>
<th>Original Image</th>
<th>Artist-1 drawing</th>
<th>Generated Caricature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 6-13(i). The second cross-validation results of artist-1.

In the first validation case (row 1), the drawing style of artist-1 does not appear to be revealed in the computer generated caricature. A careful analysis of the training set revealed that for artist-1 the style, ‘forehead elongation’, is not represented within the training set (see column 2 of figure 5-15). It should be noted here that even though the
forehead elongation style has not been captured, some facial components such as the eye, eyebrows and the mouth of the computer generated caricature resemble those components of the original caricature. However, in the second validation case (row 2), the computer generated caricature resembles the original caricature very well, especially in terms of exaggerations done to the eyes, nose and mouth.

**The Second Validation Instance of Artist-2:**

<table>
<thead>
<tr>
<th>Case</th>
<th>Original Image</th>
<th>Artist-2 drawing</th>
<th>Generated Caricature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Original Image" /></td>
<td><img src="image2.png" alt="Artist-2 drawing" /></td>
<td><img src="image3.png" alt="Generated Caricature" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4.png" alt="Original Image" /></td>
<td><img src="image5.png" alt="Artist-2 drawing" /></td>
<td><img src="image6.png" alt="Generated Caricature" /></td>
</tr>
</tbody>
</table>

Figure 6-13(ii). The second cross-validation results of artist-2.

In the first validation case (row 1), the shape of the face of the computer generated caricature is heavily elongated, which represents the unique face elongation trait of the second artist that was discussed in 6.5.2. Though the style of caricaturing used by the artist for most facial components such as ears, mouth and eyes has been captured and reproduced by the computer based scheme. However, in the second validation case (row 2), the computer generated caricature resembles the original caricature very
well, especially in terms of exaggerations done to the eyes, nose and the facial shape.

The Second Validation Instance of Artist-3:

<table>
<thead>
<tr>
<th>Case</th>
<th>Original Image</th>
<th>Artist-3 drawing</th>
<th>Generated Caricature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Original Image" /></td>
<td><img src="image2" alt="Artist-3 drawing" /></td>
<td><img src="image3" alt="Generated Caricature" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4" alt="Original Image" /></td>
<td><img src="image5" alt="Artist-3 drawing" /></td>
<td><img src="image6" alt="Generated Caricature" /></td>
</tr>
</tbody>
</table>

Figure 6-13(iii). The second cross-validation results of artist-3.

In the first validation case (row 1) of artist-3, a very similar drawing style is illustrated by both the artist's drawing and the generated caricature. Similarly, the second validation case (row 2) demonstrates that some drawing traits of artist-3 have been successfully captured, which is apparent in the exaggeration of the nose in the generated caricature. However, the amount of facial elongation of the computer generated image is more than what is expected.

The second cross-validation results presented above further demonstrate the possible use of the proposed entire face-based caricature generation approach in capturing the drawing style of an artist. Although the proposed system cannot generate caricatures
with exact resemblance to the corresponding caricatures of the artist, a significant extent of their drawing styles has been successfully captured and reproduced by the proposed technique.

6.7.4 Comparisons with Benchmark

The two validation images, 11 and 12, were finally caricatured using the Picasso algorithm of [8] for comparison with the results obtained from the proposed caricaturing system. Figure 6-14 illustrates the caricatures of the validation images produced when using the Picasso method with exaggerations to each feature performed by multiplying the original facial image's corresponding feature difference from the mean, by a factor of 1.5. It is noted that the Picasso method is a conventional linear exaggeration algorithm (see subsection 3.3.1). Whilst this approach produces a caricature, it does not produce caricatures that imitate the drawing style of a given artist, due to the fact that no drawing style capture algorithm has been used within the design.

Figure 6-14. Caricatures produced by the Picasso benchmark with $\Delta S' = 1.5 \times \Delta S$. 
6.8 Chapter Summary and Conclusions

This chapter initially provided an overview of the proposed caricature generation system with the aid of a block diagram to highlight all important steps. It further explained the relationships amongst the original image, corresponding caricature and mean face, which is widely accepted as the method used by the human psycho-visual system to capture the unique features of a facial figure. It then proceeded to propose a novel entire face-based automatic caricature generation approach, with details of neural network data preparations, settings and training procedures provided. Subsequently, a cross-validation strategy was discussed and adopted to extensively evaluate the performance of the proposed system. Experimental results obtained for all three artists were carefully analysed, which compared the drawing style of the computer generated caricatures with those of the caricatures drawn by the artists. The chapter further compared each experimental result with the corresponding drawings of the three artists, which clearly demonstrated that a significant amount of the drawing rules have been successfully captured and reproduced by the proposed entire face-based approach. Finally, well defined subjective tests were carried out to further support the claim. A detail statistical analysis of the subjective test results was carried out, which concluded that the participants have been able to correctly match the artists' caricatures with the corresponding computer generated caricatures, with a confidence level of over 99%.

This chapter proposed an entire face-based automatic caricature generation approach, which has contributed to prove the relationship concept of mean face, original image and corresponding caricature (see section 6.3). It further preliminary demonstrated that the drawing style of the artists can be captured by using neural network and is then able to create photorealistic caricatures that follow their drawing styles.
Nevertheless, potential exists to improve the accuracy of the entire face-based automatic caricature generation system by increasing the number of FDPs used in the adopted facial model. A more efficient system that is capable of more accurate caricature generation, i.e. a facial component-based approach, is presented in Chapter 7.
Chapter 7

Facial Component-Based Caricature Generation Approach

7.1 Introduction

In the entire face-based approach proposed in Chapter 6, a simple face model with 46 FDPs was used to represent both original images and caricatures. However such a low number of FDPs is practically not sufficient to represent the detail and variety of a human facial figure. Nevertheless it provides means for maintaining the computational cost of the entire face-based approach at a relatively low level. In this chapter, a novel facial component-based approach is proposed to further improve the quality of the generated caricatures, without increasing the requirement of computational resources used at a given instance of time.

This chapter proposes a novel facial component-based approach to caricature generation. It is organised as follows:

7.2 An overview

7.3 Facial component-based caricature generation approach

7.4 Experimental results and analysis

7.5 Subjective test

7.6 Further experiments

7.7 Chapter summary and conclusions
7.2 An Overview

An obvious way to improve the accuracy of the entire face-based approach proposed in Chapter 6 is to increase the number of FDPs (resolution of the caricature) that are used to represent the detail of a human face. Unfortunately this results in a major increase of the computational power and memory requirements of the automatic caricaturing algorithm per single task of training and testing. Further it is likely that caricaturists apply completely different and/or independent drawing rules to different components of a facial figure, thereby limiting the flexibility of the entire face-based approach to capture and reproduce the variability between the styles of caricaturing used by artists for different facial components.

An alternative solution to this problem is to consider the automatic caricaturing of individual facial components (such as nose, mouth, eyes etc.) with an increased total number of FDPs and then joining the components subsequently to create the caricature of the entire face. This approach is named as the facial component-based approach in this Thesis. However, training the facial components independently results in losing their relative positions on the face. As a result, a final positioning stage is required to provide a proper orientation for facial components.

In general, the component-based approach has the advantage of using an increased pixel resolution to represent each component thereby increasing the overall accuracy of caricaturing process. Further to this it has the advantage of limiting the neural network training to individual components thereby enabling the capture of often common variations of the drawing style between components of a given artist's caricature.
Due to the striking similarity of the fundamental ideas underpinning the two proposed approaches, the common procedures which have been introduced in Chapter 6 will not be discussed in the following sections.

A block diagram that summarises the structure and operation of the proposed facial component-based approach is illustrated in figure 7-1.

Figure 7-1. The proposed facial component-based automatic caricature generation algorithm.
7.3 Facial Component-Based Caricature Generation Approach

7.3.1 Preparation of Training and Validation Sets

In the facial component-based approach, the processes of training and validation set preparation remain the same as the entire face-based approach (see subsection 6.4.2) except for the fact that each facial component (i.e. eyes, eyebrows, nose, mouth, ears and face contour) is now considered as a separate subset.

In Chapter 5, two novel geometric face models were proposed. The first model, known as the simple face model, was adopted by the entire face-based approach of the Chapter 6. With a view to increase the human face representation accuracy, the second face model, i.e. the enhanced model with 143 feature points (refer to section 5.3) is adopted in the facial component-based approach. It is noted that the enhanced face model in total utilises approximately three times FDPs as the simple face model. However the number of FDPs used per single component in the component-based approach is not more than that used for the entire face in the entire face-based approach. Therefore the computational resources used per single training and testing task (e.g. mouth, nose etc., compared to the entire face) is lowered, thereby enabling any computer with medium specifications can implement and carry out the facial component-based caricaturing approach. A comparison of the two proposed face models on the basis of FDPs used per component is reproduced in table 7-1 for easy reference.
Chapter 7. Facial Component-Based Caricature Generation Approach

<table>
<thead>
<tr>
<th>Facial Component</th>
<th>Simple Face Model</th>
<th>Enhanced Face Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face Contour</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>Both Ears</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>Both Eyebrows</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>Both Eyes</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>Nose</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Mouth</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>46</strong></td>
<td><strong>143</strong></td>
</tr>
</tbody>
</table>

Table 7-1. Comparison of the two face models.

7.3.2 Neural Network Architectures

In the facial component-based approach, the architecture and parameters for constructing neural networks are identical to those that were used in the entire face-based approach and are summarised in table 6-2. However, the number of neural networks required is increased as each facial component is trained independently. Therefore, a total of twelve neural networks per artist were constructed, where the X and Y coordinates of six facial components (as listed in table 7-1) are trained separately (see section 6.5).

On top of this, for a given neural network, the number of nodes in the input and output layers, is now equal to the number of FDPs of the corresponding facial component as tabulated in table 7-2. The number of neurons in the hidden layer is designed to be same as the corresponding input/output layers (see subsection 6.7.1). Note that the left and right counterparts of facial components such as eyebrows, eyes and ears are trained together, without a significant loss in caricaturing efficiency.
Chapter 7. Facial Component-Based Caricature Generation Approach

For an easy comparison, table 7-2 summarises the neural network architectures used by the entire face-based and the facial component-based approaches.

<table>
<thead>
<tr>
<th>Neural Network Architecture</th>
<th>Entire Face-Based Approach</th>
<th>Facial Component-Based Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Neural Networks Required per Artist</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Neural Network Type</td>
<td>Feed-forward Back-propagation</td>
<td>Feed-forward Back-propagation</td>
</tr>
<tr>
<td>Number of Layers</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Number of nodes in input layer</td>
<td>46</td>
<td>No. of FDPs of component</td>
</tr>
<tr>
<td>Number of nodes in hidden layer</td>
<td>46</td>
<td>No. of FDPs of component</td>
</tr>
<tr>
<td>Number of nodes in output layer</td>
<td>46</td>
<td>No. of FDPs of component</td>
</tr>
</tbody>
</table>

Table 7-2. A comparison of neural network architectures used by the two caricaturing approaches.

7.3.3 Neural Network Training

In the facial component-based approach, the training process is identical to the entire face-based approach (see subsection 6.4.4) except all the facial components are trained separately in order to lower the computational requirements, as each facial component has a relatively low number of FDPs when compared with the FDPs used in the entire face-based caricaturing approach.

7.3.4 Neural Network Validations

Once again, a cross-validation approach is adopted to evaluate the proposed component-based caricaturing approach (see subsection 6.4.2). The validation set of each facial component is forwarded to the corresponding, trained, neural network. The validation outputs from all neural networks are collected and positioned by an additional module proposed in the next subsection 7.3.5, before passing to the final
stage, i.e. the *mesh warping* stage. The final result of the first cross-validation instance (i.e. images 11 and 12) will be fully analysed and discussed in sections 7.4 and 7.5 whilst the second instance (i.e. images 2 and 7) will be illustrated in section 7.6 for comparison purposes.

### 7.3.5 Facial Component Positioning

In the entire face-based approach proposed in Chapter 6, the advantage of training all facial components together is that the neural network is not only capable of learning the exaggerations of components but also on their relative positions. Therefore the facial components of the generated caricature are always positioned accurately relative to each other on the face. Unfortunately in the facial component-based approach, when considering components separately, the information on the relative positioning is lost. Hence overlapping of facial components can occur in the resulting caricature if the generated components are combined together without a pre-processing stage. This commonly happens in cases such as when a heavily elongated nose overlaps with the mouth or exaggerated eyes overlap with eyebrows, due to their natural closeness.

In order to remedy the above problem, the entire face-based approach is used as a means for creating a template for the exaggerated relative positioning of the facial components of the caricature. The individually caricatured facial components are then replaced the corresponding components on the template and passed to the mesh warping module for caricature generation (see subsection 6.4.5). Figure 7-2 demonstrates an example of a caricature before and after component positioning.
Figure 7-2. An example of a caricature before and after facial component positioning. (a) Original image. (b) Template (result of the entire face-based approach). (c) Overlapped components (generated from the component-based approach) without using template. (d) Positioned component according to the template.
7.4 Experimental Results and Analysis

In Chapter 6, strong experimental proof was provided as to the ability of the entire face-based approach to learn the drawing style of a caricaturist, and to subsequently produce photorealistic caricatures that are embedded with the unique drawing style of the artist. In this chapter, experiments are designed and carried out to demonstrate that the facial component-based approach is capable of providing higher accuracy results, which means that the generated caricatures have increased similarities to the corresponding artists' drawings.

Once again, MATLAB and its ANN toolbox were used as the programming language/environment [112] following the parameters provided in tables 6-2 and 7-2. Neural network training and validations as described in section 7.3 were carried out, and the experimental results of artist-1, artist-2 and artist-3 are separately presented in subsections 7.4.1, 7.4.2 and 7.4.3 respectively.
7.4.1(i) The First Validation Experiment of Artist-1:

Figure 7-3(i). (a) Original image 11 (validation case 1). (b) Caricature drawn by Artist-1. (c) Caricature generated by the entire face-based approach. (d) Caricature generated by the facial component-based approach.

Figure 7-3(i) illustrates the result of the first validation experiment of artist-1. When comparing the caricature drawn by the artist, i.e. (b), with the caricatures generated by the proposed systems, i.e. (c) and (d), most facial components of (b) look closer to those of (d) than to those of (c).

First of all, the shape of nose of (b) is more similar to that of (d) than to that of (c), as specifically indicated by the shape and size of the lower part. Besides, a more satisfactory result is obtained for the ears when using the facial component-based approach, as the shapes of ears of (b) look closer to the regularly exaggerated ears of (d) than to those of (c).

Furthermore, the sizes and shapes of the mouth, eyes and eyebrows of (b) are slightly closer to (d) than to those of (c). Finally, the shape of the face of (b) looks much similar to (d) than to that of (c) as the region of jaws shows clearly, although the overall shape of face in (d) still cannot completely capture the drawing style of (b).
7.4.1(ii) The Second Validation Experiment of Artist-1:

Figure 7-3(ii). (a) Original image t2 (validation case 2). (b) Caricature drawn by Artist-1. (c) Caricature generated by the entire face-based approach. (d) Caricature generated by the facial component-based approach.

Figure 7-3(ii) shows the second validation experiment of artist-1. When comparing the caricature drawn by the first artist, i.e. (b), with the caricatures generated by the entire face-based approach, i.e. (c), and the facial component-based approach, i.e. (d), a more satisfactory result is illustrated by (d).

Firstly, the sizes of eyes in (b) look closer to the more exaggerated eyes of (d) than to those of (c). Moreover, the shape of the face in (d) is better than in (c) as the size of forehead is more similar to that of (b), even though the forehead still cannot be fully imitated by (d). Besides, the jaws of (d) are more accurate than that of (c) when compared with (b). The reason for the above added accuracy is the increased number of FDPs used to represent both forehead and jaw areas.

Furthermore, the eyebrow distortions in (b) are more similar to that of (d) than to that of (c). Finally, the shapes of ears and nose in (b) also look slightly closer to those of (d) than to those of (c).
7.4.2(i) The First Validation Experiment of Artist-2:

![Image](a) Original image 11 (validation case 1). (b) Caricature drawn by Artist-2. (c) Caricature generated by the entire face-based approach. (d) Caricature generated by the facial component-based approach.

To further evaluate the performance of the facial component-based caricature generation algorithm, experiments similar to that carried out for the artist-1 were carried out for artist-2.

Figure 7-4(i) illustrates the result of the first validation experiment on artist-2's caricaturing products. The result clearly demonstrates that the facial component-based approach performs better than the entire face-based approach of Chapter 6. Firstly, the shapes of both ears of (b) look much closer to those of (d) than to those of (c), as demonstrated clearly by their widths. In addition, the shapes of both eyebrows and eyes in (b) are more similar to those of (d) than to those of (c).

Finally, both proposed approaches were able to generate caricatures with elongated faces, which is a dominant trait of artist-2 as discussed in Chapter 6 (see subsection 6.5.2). Although the degree of face elongations in both (c) and (d) are more than expectation, the original caricature (b) looks closer to (d) than to (c).
7.4.2(ii) The Second Validation Experiment of Artist-2:

![Figure 7-4(ii)](image)

(a) Original image 12 (validation case 2). (b) Caricature drawn by Artist-2. (c) Caricature generated by the entire face-based approach. (d) Caricature generated by the facial component-based approach.

Figure 7-4(ii) illustrates the result of the second validation experiment for artist-2's caricaturing products. It is shown that in figure (d), the ears are elongated in a balanced way, which matches well with the exaggeration of ears in (b) but not in (c). Moreover, the size of eyes in (b) looks closer to that of (d) than to that of (c). The same is observed with respect to the nose, where the degree of nose exaggeration in (b) is more similar to that of (d) than to that of (c).

Finally, the shape of the face in (b) is much similar to that of (d) than to that of (c), which is demonstrated best by exaggerations of the jaws of the facial figures. The increased number of FDPs that are used to refine the representation of jaws in the facial component-based approach can be pointed as the main reason for this improved performance.

However, a slightly tilted face and a dent at the chin of (b) still cannot be observed in (d), not in (c). Such omissions are due to the fact that these two drawing traits have been absent in the artist-2's caricatures used in the training set. Consequently, both
caricaturing approaches have not been able to imitate this drawing style. The importance of having a sufficiently large training set is reiterated from such observations.

7.4.3(i) The First Validation Experiment of Artist-3:

![Figure 7-5(i)](image)

(a) Original image 11 (validation case 1). (b) Caricature drawn by Artist-3. (c) Caricature generated by the entire face-based approach. (d) Caricature generated by the facial component-based approach.

To further evaluate the performance of the facial component-based caricature generation algorithm, experiments similar to that carried out for the artist-1 and artist-2 were carried out for artist-3.

Figure 7-5(i) illustrates the result of the first validation experiment using the caricatures of artist-3. As in the case of artist-1’s and artist-2’s caricatures, the caricature drawn by the third artist, (b), has a better correspondence to (d) than to (c), when exaggerations to the different facial components are carefully considered.
Firstly, the contour of the face in (b) is more similar to that of (d) than to that of (c). It is evident that the greatly increased number of FDPs in the facial component-based caricature generation approach has significantly improved the prediction accuracy by providing a more detailed description to the face contour. Besides, the style of eye and eyebrow modifications in (b) is closer to that of (d) than to that of (c), as both components are precisely exaggerated in (d). Unfortunately, the width of the face and both ears in (d) are still slightly shorter than those in (b). After all, completely capture and imitate the drawing style of an artist is an extremely difficult task.

7.4.3(ii) The Second Validation Experiment of Artist-3:

(a) (b) (c) (d)

Figure 7-5(ii). (a) Original image 12 (validation case 2). (b) Caricature drawn by Artist-3. (c) Caricature generated by the entire face-based approach. (d) Caricature generated by the facial component-based approach.

Finally the figure 7-5(ii) illustrates the result of the second validation experiment for the caricatures of artist-3, which further demonstrates that the proposed facial component-based approach provides a higher quality caricature than the entire face-based approach of Chapter 6.

Although the caricatures generated by the two automatic caricature generation approaches, (c) and (d), share a large amount of similarity, the caricature produced by
the component-based approach shows better resemblance to the caricature drawn by the artist. First of all, both size and shape of eye exaggerations in (b) are much similar to that of (d) than to that of (c). Moreover, the shape of the jaws of the original caricature, (b), drawn by the artist, is accurately predicted by the facial component-based approach, (d), which has adopted the “Enhanced Face Model” (see section 5.3) with three times more feature points than the “Simple Face Model” used by the entire face-based approach. Finally, the exaggerations made to both ears and eyebrows of (b) look slightly closer to those of (d) than to those of (c). As a result, the collective visual effect of facial components in (d), illustrates an improved resemblance to the caricature drawn by the artist-3, i.e. (b).
Chapter 7. Facial Component-Based Caricature Generation Approach

7.5 Subjective Test

As the esthetic analysis and justification of the photorealistic nature of generated caricatures by few individuals (e.g. author and the supervisory team) in section 7.4 can be subjective, a further scientific subjective test that involves judgments from a number of participants was carried out to further confirm the validity of conclusions made by the above validation experiments.

In this subjective test, a different questionnaire was designed to compare the effectiveness of the entire face-based and facial component-based approaches. For each caricaturist and for each validation image, 46 subjects were provided with the original image, the caricature drawn by the artist and the photorealistic caricatures generated from the two automatic caricature generation approaches. The computer generated caricatures were presented in a mixed manner, without stating the approach that was followed to obtain it, i.e. whether the entire or component-based approach was used. The subjects were asked to decide which of the two computer generated images was more similar to the caricature drawn by the artist.

This questionnaire gives a more direct comparison between the subjective quality of the two approaches, as compared to repeating the subjective test of Chapter 6 on the component-based approach and subsequent comparison. Figure 7-6 (i) – (iii) lists and illustrates all key questions used in the test. Participants were invited to choose the computer generated caricature that resembles the closest drawing style to the artist’s drawing (column 2), for each of the two validation cases. Note that the computer generated caricatures are illustrated in mixed/random order, in columns 3 and 4. A complete questionnaire can be found in Appendix C.
Chapter 7. Facial Component-Based Caricature Generation Approach

**Question 1:**

<table>
<thead>
<tr>
<th>Case</th>
<th>Original image</th>
<th>Artist-1 drawing</th>
<th>Computer Generated Caricatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Original Image" /></td>
<td><img src="image2" alt="Artist-1 Drawing" /></td>
<td><img src="image3" alt="Computer Generated Caricatures" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4" alt="Original Image" /></td>
<td><img src="image5" alt="Artist-1 Drawing" /></td>
<td><img src="image6" alt="Computer Generated Caricatures" /></td>
</tr>
</tbody>
</table>

Answer:

Figure 7-6(i). Subjective test on artist-1’s computer generated caricatures.

**Question 2:**

<table>
<thead>
<tr>
<th>Case</th>
<th>Original image</th>
<th>Artist-2 drawing</th>
<th>Computer Generated Caricatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image7" alt="Original Image" /></td>
<td><img src="image8" alt="Artist-2 Drawing" /></td>
<td><img src="image9" alt="Computer Generated Caricatures" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image10" alt="Original Image" /></td>
<td><img src="image11" alt="Artist-2 Drawing" /></td>
<td><img src="image12" alt="Computer Generated Caricatures" /></td>
</tr>
</tbody>
</table>

Answer:

Figure 7-6(ii). Subjective test on artist-2’s computer generated caricatures.
Question 3:

<table>
<thead>
<tr>
<th>Case</th>
<th>Original image</th>
<th>Artist-3 drawing</th>
<th>Computer Generated Caricatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Original image" /></td>
<td><img src="image2" alt="Artist-3 drawing" /></td>
<td><img src="image3" alt="Computer Generated Caricatures" /></td>
</tr>
<tr>
<td></td>
<td><img src="answer1" alt="Answer" /></td>
<td><img src="answer1" alt="Answer" /></td>
<td><img src="answer1" alt="Answer" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4" alt="Original image" /></td>
<td><img src="image5" alt="Artist-3 drawing" /></td>
<td><img src="image6" alt="Computer Generated Caricatures" /></td>
</tr>
<tr>
<td></td>
<td><img src="answer1" alt="Answer" /></td>
<td><img src="answer1" alt="Answer" /></td>
<td><img src="answer1" alt="Answer" /></td>
</tr>
</tbody>
</table>

Figure 7-6(iii). Subjective test on artist-3’s computer generated caricatures.

The results of the subjective test are presented in figure 7-7. The bar chart represents the percentage of participants who found that the facial component-based approach was better than the entire faced-based approach. For all three artists, more than 70% of the subjects answered that the facial component-based approach provided more superior quality caricatures than the entire face-based approach in each validation case. Therefore it can be concluded that the facial component-based approach has a better ability to capture the drawing style of an individual caricaturist as compared to the entire face-based approach. This further justifies the claims made in section 7.2.
Chapter 7. Facial Component-Based Caricature Generation Approach

Figure 7-7. Results of the subjective test that compares caricatures generated using the two approaches.

The subjective test results obtained above were further analysed using statistical decision theory [119], as described in Chapter 6. The details of this analysis can be presented as follows:

If $p$ is the probability of participants choosing the caricatures generated by the facial component-based approach rather than the entire face-based approach, then one of the following two hypotheses has to be decided:

$H_0$: $p = 0.5$, and participants randomly selected a caricature from the two provided, i.e. results were due to chance.

$H_1$: $p > 0.5$, and participants chose the caricatures generated by the facial component-based approach instead of the entire face-based approach deliberately, i.e. the facial component-based approach performed better than the entire face-based approach.

If the hypothesis $H_0$ is true, the mean, $\mu$, and standard deviation, $\sigma$, of the number of
participants chose the caricatures generated by the facial component-based approach is given by

\[ \mu = Np = 46(0.5) = 23 \]

\[ \sigma = \sqrt{Npq} = \sqrt{46(0.5)(0.5)} = 3.39 \]

where \( N \) is the sample size, \( p \) is the population proportion of successes and \( q \) is equal to \( 1-p \).

Once again, a one-tailed test with a significance level of 0.01 is adopted, as the objective of this subjective test is to investigate the number of participants that chose the caricatures generated by the facial component-based approach, but not the entire face-based approach. For a one-tailed test at a level of significance of 0.01, the critical value of \( z \) is 2.33. Thus the decision rule or test of significance is:

1. If the z score observed is greater than 2.33, the results are significant at the 0.01 level and participants chose the facial component-based approach generated caricatures instead of the entire face-based approach deliberately.

2. If the z score is less than 2.33 the results are due to chance, i.e. not significant at the 0.01 level.

By substituting the number of participants who chose the caricature generated by the facial component-based approach, \( S \), for each validation case, into equation (11) of Chapter 6, their z scores are calculated. They are tabulated in table 7-3.
Table 7-3. z score calculations of the participants who chose the caricatures generated by the facial component-based approach.

As all the z scores are greater than 2.33, decision (1) holds in all cases, i.e. it can be concluded at the .01 level that participants chose the caricatures generated by the facial component-based approach instead of the entire face-based approach deliberately. As a result, the hypothesis that “the facial component-based approach performed better than the entire face-based approach”, was found to hold at a confidence level over 99%, for both validation cases of all three artists.
7.6 Further Experiments

The second instance of cross-validation results for all three artists is illustrated in figure 7-8. The same methodology as the first instance was carried out (see section 7.3) except that the original images 2 and 7 of figure 5-15 were used for validations, whereas images 11 and 12 of figure 5-15 were included in the training set. Note that in figure 7-8, cases 1 and 2 represent validation experiments of original images 2 and 7 respectively.

The Second Cross-Validation Experiments of Artist-1:

<table>
<thead>
<tr>
<th>Case</th>
<th>Original Image</th>
<th>Artist-1 drawing</th>
<th>Entire face-based</th>
<th>Component-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="a" alt="Original Image" /></td>
<td><img src="b" alt="Artist-1 drawing" /></td>
<td><img src="c" alt="Entire face-based" /></td>
<td><img src="d" alt="Component-based" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="e" alt="Original Image" /></td>
<td><img src="f" alt="Artist-1 drawing" /></td>
<td><img src="g" alt="Entire face-based" /></td>
<td><img src="h" alt="Component-based" /></td>
</tr>
</tbody>
</table>

Figure 7-8(i). The second cross-validation results of artist-1.

In the first validation case (row 1), although both proposed approaches cannot precisely predict the drawing style of artist-1 as the trait of forehead elongation in (b) is absent from the training set (see column 2 of figure 5-15), the component-based
approach shows an improvement to the entire face-based approach when comparing both computer generated caricatures, i.e. (c) and (d), to artist's drawing (b). This is demonstrated by the region of jaws and size of ears. Similarly in the second validation case, the heavily exaggerated eyes and nose in (f) look closer to (h) than to those of (g). Even though the exaggeration of ears by the component-based approach (h) still cannot capture the corresponding modification made by artist-1 in (f), it is better than (g).

The Second Cross-Validation Experiments of Artist-2:

<table>
<thead>
<tr>
<th>Case</th>
<th>Original Image</th>
<th>Artist-2 drawing</th>
<th>Entire face-based</th>
<th>Component-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
<td>(d)</td>
</tr>
<tr>
<td>2</td>
<td>(e)</td>
<td>(f)</td>
<td>(g)</td>
<td>(h)</td>
</tr>
</tbody>
</table>

Figure 7-8(ii). The second cross-validation results of artist-2.

In both validation cases (rows 1 and 2), the caricatures generated by the component-based approach, (d) and (h), slightly improved the entire face-based approach, i.e. (c) and (g), when compared with the corresponding artist-2's drawings.
(b) and (f). In the first validation case (row 1), the improvements appear in the shape of face and ears in (d). Likewise, in the second validation case (row 2), the more exaggerated eyes and balanced ears in (h) demonstrate a better ability of drawing style prediction than in (g).

The Second Cross-Validation Experiments of Artist-3:

<table>
<thead>
<tr>
<th>Case</th>
<th>Original Image</th>
<th>Artist-3 drawing</th>
<th>Entire face-based</th>
<th>Component-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
<td>(d)</td>
</tr>
<tr>
<td>2</td>
<td>(e)</td>
<td>(f)</td>
<td>(g)</td>
<td>(h)</td>
</tr>
</tbody>
</table>

Figure 7-8(iii). The second cross-validation results of artist-3.

It can be seen that the facial component-based approach provides satisfactory results in both validation experiments of artist-3. In the first validation case (row 1), the modifications of eyebrows and jaw region in artist-3's drawing, i.e. (b), are more similar to those of (d) than to those of (c). In the second validation case (row 2), even though both proposed approaches over-elongated the faces in the generated caricatures, i.e. (g) and (h), when compared to artist-3's drawing (f), the amount of exaggeration in (h) demonstrates an improvement to (g).
7.7 Chapter Summary and Conclusions

This chapter proposed a novel facial component-based caricature generation approach, which is capable of further improving the quality of the caricatures generated by the entire face-based approach of Chapter 6. Initially the chapter provided an overview of the proposed algorithm with a block diagram to illustrate the system architecture. It then proceeded to propose the novel facial component-based caricature generation approach. The differences between the neural network architectures used in the facial component-based and entire face-based approaches were discussed. It further proposed a facial component positioning module to solve the overlapping problem of the generated components. Subsequently, the experimental results obtained from the facial component-based approach, when using the caricatures of all three artists, were carefully compared with those produced by the entire face-based approach of Chapter 6. This in-depth analysis clearly demonstrated that the facial component-based approach achieved higher drawing style capture accuracy than the entire face-based approach. Finally, a subjective test, which aimed at investigating which proposed approach (entire face-based vs. facial component-based) generated better results, was carried out to support the claim further. A detailed statistical analysis of the results was carried out, which concluded that, “the facial component-based approach performed better than the entire face-based approach”, with a confidence level of over 99%.

This chapter proposed a facial component-based caricature generation approach, which has contributed to produce higher quality caricatures than the entire face-based approach. The increased number of FDPs used within the facial component based approach and the individual component based training procedure adopted that allowed the capture of variability in styles are the fundamental reasons for this enhanced
performance over the entire face based approach presented in chapter 6. Further the computational resources utilised per component based caricaturing task of the component based approach is lower than the computational resources that needs to be utilised for the whole face based caricaturing task of the entire face-based approach. Moreover, the addition of the facial component positioning module has also contributed to tackle the problem of overlapping between generated components.

Although the proposed approaches, at their present stage of development, are unable to completely predict and reproduce the exact drawing style adopted by an artist, they provide a significant move towards this ultimate goal. Better results could have been obtained by training the neural networks with a larger number of samples, from a given artist.
Chapter 8

Thesis Summary, Conclusions and Future Work

8.1 Introduction
This chapter provides a summary of the research presented in this Thesis and draws upon overall conclusions from the experimental results and analysis presented in the individual chapters. It further discusses the practical challenges, limitations, potential applications and future work related to the automatic caricature generation approaches proposed in this Thesis.

The chapter is organised as follows:

8.2 Thesis summary
8.3 Practical challenges
8.4 Applications
8.5 Limitations of research and future work
8.6 Thesis conclusions
8.2 Thesis Summary

The aim of this research was to develop a system by which caricatures, embedded with the drawing style of a given artist, can be automatically generated. The main contributions of the research presented in this Thesis are:

1. A comprehensive review of the fundamental concepts of drawing caricature and a complete survey of existing automatic caricature generation technologies.

2. A novel face modelling framework, which can provide robust platforms for facial caricaturing.

3. A novel, neural network based automatic caricature generating approach that is capable of capturing and subsequently reproducing the drawing style of an artist based on the entire face (i.e. global facial features) of caricatures.

4. A novel, neural network based automatic caricature generating approach that is capable of capturing and subsequently reproducing the drawing style of an artist based on the individual facial components (e.g. eyes, nose, mouth etc.) of caricatures.

The above contributions were presented within the Thesis as follows:

Part 1: Introduction, Review and Background Technologies

Chapter 1 provided an overview of the project, a problem description, research motivation and objectives. It further outlined the methodology, approach and processes proposed within the Thesis to tackle the research problem.

Chapter 2 gave an introduction to caricature, including its definition, history and the
applications of caricature in our daily life. It further discussed the procedures adopted to draw caricatures by both professional artists and beginners.

Chapter 3 provided a comprehensive review of the existing computer based caricaturing systems. The strengths and weaknesses of each approach were critically compared and their common problems were discussed.

Chapter 4 provided an introduction to the background technologies used within the project, which included the state-of-the-art geometric face models, artificial intelligence technology and the fundamental image processing techniques that are adopted by the proposed caricaturing system.

Part 2: Contributions of Research

Chapter 5 proposed a novel face modelling framework with two different geometric face models, which provided a platform to the face modelling stage, which is a part of the proposed drawing style capture algorithms.

Chapter 6 proposed a novel entire face-based caricature generation approach to capture the drawing style of an artist. The experimental results demonstrated that the adopted neural networks were capable of accomplishing the drawing style capturing task and subsequently creating satisfactory photorealistic caricatures.

Chapter 7 proposed a novel facial component-based caricature generation approach, which demonstrated that it has the capability of improving the accuracy of drawing style capture and prediction whilst maintaining the computational resource requirements per training task at a level below that of the entire face-based approach.
8.3 Practical Challenges

The experimental results and analysis provided in Chapters 6 and 7 demonstrated that the proposed example-based neural network approaches to automatic caricature generation were able to capture, learn and re-produce the drawing style of a caricaturist. Though these proofs were based on testing a limited set of original image-caricature pairs drawn by a limited number of caricaturists, the author strongly believes that the significantly good experimental results obtained (both of an objective and subjective nature) and the detailed analysis carried out justify the claims made.

Training on a larger set of original image-caricature pairs (belonging to different people but caricatured by the same artist) will help improve prediction further. Unfortunately, requesting a caricaturist to draw a large number of caricatures of a large database of facial images is time consuming and somewhat unreasonable. This is a major practical challenge. Besides, the size of the training set is another considerable issue when processing and storage costs are considered. Therefore, capturing the drawing style of an artist from a limited dataset and still obtaining satisfactory result is a key challenge of the project. Added to this, a large number of FDPs have to be defined if one is to expect a higher quality of caricature generation. However this increases the need of computational resources for pre-processing.

After careful consideration of the above mentioned practical limitations of the experimentation process, a cross validation scheme for analysing the performance of the proposed system was adopted. The use of a facial component-based approach was further proposed in which a larger amount of FDPs can be defined on a per-component basis, thus maintaining the cost of training per-component below that of the entire face-based approach.
8.4 Applications

Computer based caricature generation has been applied in face recognition and psychological studies as it is capable of automatically enhancing the distinctive facial features of a particular face, which provides a direct manner for testing the role of distinctiveness in face perception and recognition [24,26,31]. Psychological studies suggest that faces rated as distinct by viewers are better recognised than faces rated as typical [30]. Thus researchers have made use of caricatures, veridical images and anti-caricatures (see subsection 3.2.1) to examine the speed and accuracy of face recognition by presenting them to observers. Experimental results have demonstrated that caricatures of faces are recognised more quickly and accurately than veridical images of faces [24-26]. Further studies verified that caricature has a “better likeness” to individual than its veridical images [31]. Therefore, the caricature generation approaches proposed in this Thesis could significantly extend the practical applicability of the above research as the proposed systems are able to automatically generating caricatures of a subject with an embedded drawing style of an expert caricaturist. As the individuals illustrated by caricatures of professional artists can be easily identified by general public, a face recognition system that adopts the proposed automatic caricature generation algorithms that is trained to reproduce caricatures with an embedded expert-style, will have a better chance of success as compared to adopting a computer exaggerated caricature generation algorithm.

The proposed scheme has further potential to be applied in many practical application areas apart from for pure entertainment purposes and face recognition. Though generating a caricature from a facial image is considered a popular art, the ability of a computer based system to learn the drawing style of an artist plays a significant role in its possible application domains. One example is the use of the proposed approach
to reproduce caricatures that have embedded styles of late, historically famous caricaturists. This may allow their work to achieve an eternal status. Further a busy caricaturist can make use of the proposed system to provide assistance in generating copyrighted computer generated caricatures, embedding with his/her drawing style. Indirectly the proposed drawing style capture method can be used in more advanced computer vision applications, e.g. copyright theft prevention, assistance during legal proceedings, person identification via police drawings, etc.
8.5 Limitations of Research and Further Work

The evaluation of the performances of the automatic caricaturing systems proposed in this Thesis has a number of limitations, due to the constraints of test data (i.e. original facial figures, caricatures drawn by an artist etc).

The number of artists used in the experiments and the number of caricatures from each artist used in the training were limited. Despite these limitations, the experimental results supported by the detailed analysis have clearly demonstrated the effectiveness of the proposed algorithms.

A further limitation of the proposed algorithms is the manual marking of feature points. This makes the proposed algorithms to be more accurately classified as semi-automatic, in its present state of implementation, rather than to be classified as fully automatic. The manual marking strategy was adopted due to the need of simplifying the implementation of the algorithm as the focus of research was on the efficient capture of the drawing style, rather than the flexibility of implementation. However computer vision techniques adopted by the MPEG-4 FDP marking techniques [98-102] can be straightforwardly used here to resolve this limitation.

The neural networks used within the proposed research utilised a single hidden layer architecture. The use of multiple hidden layers can be investigated. To this effect, preliminary experiments using MATLAB’s ANN toolbox were not successful due to a suspected implementation problem of the MATLAB’s associated library functions. Unfortunately, implementation of the above using an alternative programming language was outside the time constraints of the present research.
Moreover, the caricatures generated by the proposed system are photorealistic instead of line-drawing. According to the literature review in Chapter 3, nearly half of the existing caricature generation systems adopted the photorealistic approach [3,9,12-14] while the rest adopted the line-drawing [2,4,10-11,15-17], as both approaches have their advantages and disadvantages. However, in this particular project, the line-drawing approach would provide an easier and more direct comparison between artists’ drawings with the system outputs, especially for users who are novice to caricaturing. Therefore, the proposed system can be further improved by converting the photorealistic results into line-drawings [61-62], which will help increase the accuracy of system analysis and evaluation.

Further the original images used in caricaturing were 2D frontal views of a facial figure with no-pose variation. Furthermore all selected facial figures for caricaturing were male, had short hair and did not wear accessories. All these constraints contributed to the simplicity of the present implementation of the proposed automatic caricature generation system, thereby enabling the research to concentrate on more important, fundamental issue of drawing style capture and reproduction. In the future, the proposed systems and their evaluations can be improved via attempts to caricature faces of different genders, ages, ethnic origins, facial expressions, hair styles, accessories, etc. Developing entire human body/figure caricaturing can also be considered as against facial only caricaturing. Capturing the drawing style of artists belonging to different cultures, religious backgrounds etc, and studying the details of their styles can have many applications in psychology, sociology, science and education. Recreating caricatures with these culturally different styles will have applications in entertainment, art and cinema.
Further to the above mentioned, short-term, research ideas, more advanced research of a longer term nature can be carried out. The outputs of the trained neural network used in the present approach can be further processed by artificial intelligence technology to extract linguistic rules [122-123], which can be used to present the captured drawing rules in a human understandable manner. Similarly, a linguistic decision tree approach can be used to replace the neural network for the purpose of investigating the drawing style of an artist. Both of these suggestions will help in the understanding and drawing style comparisons of different artists in both, artistic and psychological ways. Furthermore, inviting an artist to draw caricatures of the same subject from different angles would help construct a corresponding 3D caricature, which makes capturing and imitating the drawing style of an artist in a 3D domain possible. Finally, the role of mean face can be further investigated by replacing it with one calculated from a specific ethnic origin. It is likely that using a mean face generated from a particular ethnic origin, in the automatic caricaturing of the work of an artist from the same ethnic origin, using the proposed approaches, would enable more accurate identification of drawing traits unique to artists of that culture. Further, examples such as caricaturing a European face with a Chinese mean face, can be investigated.

In the recent past, Support Vector Machines (SVMs), have gained popularity as a replacement to neural networks, particularly in cases where the training dataset is limited in size [124-126]. The use of SVMs to replace the neural network used within the present context of this research can be investigated, and has potential to provide more accurate results.
8.6 Thesis Conclusions

The problem statement of the proposed research that was formulated in Chapter 1 can be revisited as follows:

*How is it possible to capture the drawing style of an artist and automatically generate caricatures by a computer, depicting the captured drawing style?*

This Thesis identified the main drawback of existing automatic caricature generation systems in their inability to capture and re-produce the drawing style of a given artist. Addressing this shortcoming, two novel example-based neural network approaches that were able to capture and identify the drawing style of an artist by training a neural network on original images and corresponding caricatures drawn by an artist were proposed. The trained neural network was subsequently used to automatically generate complete face caricatures possessing the drawing style of the artist. Subjective and analytical experimental results were provided to demonstrate the possible effective use of the algorithms in automatic caricature generation.

In conclusion it can be stated that, this work is the first neural network based caricature generation system that has a proven ability to identify the drawing style of a given artist and reproduce caricatures with the style embedded. To this effect the research is novel and contributes significantly towards the advancement of the research area of automatic caricature generation.
REFERENCES


[53] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative


[56] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer
version. Technical report, University of Manchester, Manchester, United
Kingdom, 2000.

sketch generation with non-parametric sampling. In Eighth International


[59] T. Fujiwara, H. Koshimizu, K. Fujimura, H. Kihara, Y. Noguchi and N.
Ishikawa. 3D Modelling System of Human Face and Full 3D Facial
Caricaturing. In The Third International Conference on 3-D Digital Imaging

expressions using 2D facial image. In The Fourth Annual ACIS
International Conference on Computer and Information Science, pages
228-232, 2005.

[61] Thinker Software. Photo to Sketch (version 4.2).


### Appendix A - Specifications of Facial Definition Parameters

<table>
<thead>
<tr>
<th>Facial Component</th>
<th>FDP</th>
<th>Text description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lips</strong></td>
<td>2.2</td>
<td>Middle point of inner upper lip contour</td>
</tr>
<tr>
<td></td>
<td>2.3</td>
<td>Middle point of inner lower lip contour</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>Left corner of inner lip contour</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>Right corner of inner lip contour</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>Midpoint between FDPs 2.2 and 2.4</td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td>Midpoint between FDPs 2.2 and 2.5</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>Midpoint between FDPs 2.3 and 2.4</td>
</tr>
<tr>
<td></td>
<td>2.9</td>
<td>Midpoint between FDPs 2.3 and 2.5</td>
</tr>
<tr>
<td><strong>Chin and Jaws</strong></td>
<td>2.1</td>
<td>Bottom of the chin</td>
</tr>
<tr>
<td></td>
<td>2.10</td>
<td>Chin boss</td>
</tr>
<tr>
<td></td>
<td>2.11</td>
<td>Chin left corner</td>
</tr>
<tr>
<td></td>
<td>2.12</td>
<td>Chin right corner</td>
</tr>
<tr>
<td></td>
<td>2.13</td>
<td>Left corner of jaw bone</td>
</tr>
<tr>
<td></td>
<td>2.14</td>
<td>Right corner of jaw bone</td>
</tr>
<tr>
<td><strong>Eyes</strong></td>
<td>3.1</td>
<td>Centre of upper inner left eyelid</td>
</tr>
<tr>
<td></td>
<td>3.2</td>
<td>Centre of upper inner right eyelid</td>
</tr>
<tr>
<td></td>
<td>3.3</td>
<td>Centre of lower inner left eyelid</td>
</tr>
<tr>
<td></td>
<td>3.4</td>
<td>Centre of lower inner right eyelid</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td>Centre of the pupil of left eye</td>
</tr>
<tr>
<td></td>
<td>3.6</td>
<td>Centre of the pupil of right eye</td>
</tr>
<tr>
<td></td>
<td>3.7</td>
<td>Left corner of left eye</td>
</tr>
<tr>
<td></td>
<td>3.8</td>
<td>Left corner of right eye</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
<td>Centre of lower outer left eyelid</td>
</tr>
<tr>
<td></td>
<td>3.10</td>
<td>Centre of lower outer right eyelid</td>
</tr>
<tr>
<td></td>
<td>3.11</td>
<td>Right corner of left eye</td>
</tr>
<tr>
<td></td>
<td>3.12</td>
<td>Right corner of right eye</td>
</tr>
<tr>
<td></td>
<td>3.13</td>
<td>Centre of upper outer left eyelid</td>
</tr>
<tr>
<td></td>
<td>3.14</td>
<td>Centre of upper outer right eyelid</td>
</tr>
<tr>
<td><strong>Eyebrows</strong></td>
<td>4.1</td>
<td>Right corner of left eyebrow</td>
</tr>
<tr>
<td></td>
<td>4.2</td>
<td>Left corner of right eyebrow</td>
</tr>
<tr>
<td></td>
<td>4.3</td>
<td>Uppermost point of the left eyebrow</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>Uppermost point of the right eyebrow</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>Left corner of left eyebrow</td>
</tr>
<tr>
<td></td>
<td>4.6</td>
<td>Right corner of right eyebrow</td>
</tr>
<tr>
<td><strong>Cheeks</strong></td>
<td>5.1</td>
<td>Centre of the left cheek</td>
</tr>
<tr>
<td></td>
<td>5.2</td>
<td>Centre of the right cheek</td>
</tr>
<tr>
<td></td>
<td>5.3</td>
<td>Left cheek bone</td>
</tr>
<tr>
<td></td>
<td>5.4</td>
<td>Right cheek bone</td>
</tr>
<tr>
<td><strong>Tongue</strong></td>
<td>6.1</td>
<td>Tip of the tongue</td>
</tr>
<tr>
<td></td>
<td>6.2</td>
<td>Centre of the tongue body</td>
</tr>
<tr>
<td></td>
<td>6.3</td>
<td>Left border of the tongue</td>
</tr>
<tr>
<td></td>
<td>6.4</td>
<td>Right border of the tongue</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td><strong>Mouth Contour</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>Middle point of outer upper lip contour</td>
<td></td>
</tr>
<tr>
<td>8.2</td>
<td>Middle point of outer lower lip contour</td>
<td></td>
</tr>
<tr>
<td>8.3</td>
<td>Left corner of outer lip contour</td>
<td></td>
</tr>
<tr>
<td>8.4</td>
<td>Right corner of outer lip contour</td>
<td></td>
</tr>
<tr>
<td>8.5</td>
<td>Midpoint between FOPs 8.3 and 8.1</td>
<td></td>
</tr>
<tr>
<td>8.6</td>
<td>Midpoint between FOPs 8.4 and 8.1</td>
<td></td>
</tr>
<tr>
<td>8.7</td>
<td>Midpoint between FOPs 8.3 and 8.2</td>
<td></td>
</tr>
<tr>
<td>8.8</td>
<td>Midpoint between FOPs 8.4 and 8.2</td>
<td></td>
</tr>
<tr>
<td>8.9</td>
<td>Right high point of Cupid's bow</td>
<td></td>
</tr>
<tr>
<td>8.10</td>
<td>Left high point of Cupid's bow</td>
<td></td>
</tr>
<tr>
<td><strong>Nose</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.1</td>
<td>Left nostril border</td>
<td></td>
</tr>
<tr>
<td>9.2</td>
<td>Right nostril border</td>
<td></td>
</tr>
<tr>
<td>9.3</td>
<td>Nose tip</td>
<td></td>
</tr>
<tr>
<td>9.4</td>
<td>Bottom right edge of nose</td>
<td></td>
</tr>
<tr>
<td>9.5</td>
<td>Bottom left edge of nose</td>
<td></td>
</tr>
<tr>
<td>9.6</td>
<td>Right upper edge of nose bone</td>
<td></td>
</tr>
<tr>
<td>9.7</td>
<td>Left upper edge of nose bone</td>
<td></td>
</tr>
<tr>
<td>9.12</td>
<td>Middle lower edge of nose bone (or nose bump)</td>
<td></td>
</tr>
<tr>
<td>9.13</td>
<td>Left lower edge of nose bone</td>
<td></td>
</tr>
<tr>
<td>9.14</td>
<td>Right lower edge of nose bone</td>
<td></td>
</tr>
<tr>
<td>9.15</td>
<td>Bottom middle edge of nose</td>
<td></td>
</tr>
<tr>
<td><strong>Teeth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.8</td>
<td>Top of the upper teeth</td>
<td></td>
</tr>
<tr>
<td>9.9</td>
<td>Bottom of the lower teeth</td>
<td></td>
</tr>
<tr>
<td>9.10</td>
<td>Bottom of the upper teeth</td>
<td></td>
</tr>
<tr>
<td>9.11</td>
<td>Top of the lower teeth</td>
<td></td>
</tr>
<tr>
<td><strong>Ears</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.1</td>
<td>Top of left ear</td>
<td></td>
</tr>
<tr>
<td>10.2</td>
<td>Top of right ear</td>
<td></td>
</tr>
<tr>
<td>10.3</td>
<td>Back of left ear</td>
<td></td>
</tr>
<tr>
<td>10.4</td>
<td>Back of right ear</td>
<td></td>
</tr>
<tr>
<td>10.5</td>
<td>Bottom of left ear lobe z</td>
<td></td>
</tr>
<tr>
<td>10.6</td>
<td>Bottom of right ear lobe</td>
<td></td>
</tr>
<tr>
<td>10.7</td>
<td>Lower contact point between left lobe and face</td>
<td></td>
</tr>
<tr>
<td>10.8</td>
<td>Lower contact point between right lobe and face</td>
<td></td>
</tr>
<tr>
<td>10.9</td>
<td>Upper contact point between left ear and face</td>
<td></td>
</tr>
<tr>
<td>10.10</td>
<td>Upper contact point between right ear and face</td>
<td></td>
</tr>
<tr>
<td><strong>Head and Hair</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>Top of spine (centre of head rotation)</td>
<td></td>
</tr>
<tr>
<td>11.1</td>
<td>Middle border between hair and forehead</td>
<td></td>
</tr>
<tr>
<td>11.2</td>
<td>Right border between hair and forehead</td>
<td></td>
</tr>
<tr>
<td>11.3</td>
<td>Left border between hair and forehead</td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td>Hairline</td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td>Hair thickness over FDP 11.4</td>
<td></td>
</tr>
<tr>
<td>11.6</td>
<td>Back of skull</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B - Subjective Test Questionnaire 1

The objective of this questionnaire is to investigate the drawing styles of different artists.

In the following table, the first column is the facial images of ten different people, each row represents a person. Three professional caricaturists are invited to draw corresponding caricatures for each facial image. Their products are shown in columns two, three and four accordingly.

Different artists have different drawing styles, please look at their drawings and recognise their specific styles carefully before answering the questions below.

<table>
<thead>
<tr>
<th>Original Images</th>
<th>Drawings of the 1st Artist</th>
<th>Drawings of the 2nd Artist</th>
<th>Drawings of the 3rd Artist</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Image 1" /></td>
<td><img src="image2" alt="Image 2" /></td>
<td><img src="image3" alt="Image 3" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4" alt="Image 4" /></td>
<td><img src="image5" alt="Image 5" /></td>
<td><img src="image6" alt="Image 6" /></td>
</tr>
</tbody>
</table>
Question 1:

In the following table, the first row shows the 11th original image and its corresponding caricatures drawn by the artists. The second row is caricatures generated by the proposed system after learning from the above drawings, and they are illustrated in random order. Please match each artist drawing to a computer generated image with the closest style by filling the answer at the bottom. Note that both hair and neck are not considered as the style of an artist in the project.

<table>
<thead>
<tr>
<th>Original Images</th>
<th>Drawings of the 1st Artist</th>
<th>Drawings of the 2nd Artist</th>
<th>Drawings of the 3rd Artist</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
</tr>
</tbody>
</table>

Computer Generated Caricatures (in random order)

Answer:
Question 2:

Likewise, the first row of the following table shows the 12th original image and its corresponding caricatures drawn by the artists. Once again, please match the artists’ drawings with the computer generated caricatures in the second row.

<table>
<thead>
<tr>
<th></th>
<th>Original Images</th>
<th>Drawings of the 1st Artist</th>
<th>Drawings of the 2nd Artist</th>
<th>Drawings of the 3rd Artist</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td><img src="image1" alt="Original Image" /></td>
<td><img src="image2" alt="Drawing 1" /></td>
<td><img src="image3" alt="Drawing 2" /></td>
<td><img src="image4" alt="Drawing 3" /></td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Computer Generated Caricatures (in random order)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image5" alt="Caricature 1" /></td>
</tr>
</tbody>
</table>

Answer:
Thank you very much for your participation. If you have any questions, feel free to contact:

Ka Hang Lai  
Research School of Informatics  
Department of Computer Science  
Holywell Park  
Loughborough University  
Loughborough LE11 3TU  
United Kingdom  

Telephone: +44 (0)1509 635721  
Email: K.H.Lai@lboro.ac.uk
Appendix C - Subjective Test Questionnaire 2

Two different approaches to caricature generation have been proposed. The objective of this questionnaire is to investigate which approach, the Entire Face-Based (EFB) or the Facial Component-Based (FCB), generates better results.

In each of the following questions, the first column of the table shows two original images. The second column is their corresponding caricatures drawn by an artist. The third and the fourth column are their corresponding caricatures generated by the proposed system based on two different approaches, as they are mixed up with each other.

Please compare the facial components, such as eyebrows, eyes, ears, nose, mouth and face, of the third and the fourth column in each row carefully, and give a tick underneath to indicate the one that looks closer to the caricature drawn by the artist in the second column.

Thank you very much for your participation. If you have any questions, feel free to contact:

Ka Hang Lai
Research School of Informatics
Department of Computer Science
Holywell Park
Loughborough University
Loughborough LE11 3TU
United Kingdom

Telephone: +44 (0)1509 635721
Email: K.H.Lai@lboro.ac.uk
Question 1:

<table>
<thead>
<tr>
<th>Original Images</th>
<th>Drawings of the 1st Artist</th>
<th>Computer Generated Caricatures (in random order)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image11.png" alt="Image 11" /></td>
<td><img src="image11A.png" alt="Image 11A" /></td>
<td><img src="image11B.png" alt="Image 11B" /></td>
</tr>
<tr>
<td><img src="image12.png" alt="Image 12" /></td>
<td><img src="image12A.png" alt="Image 12A" /></td>
<td><img src="image12B.png" alt="Image 12B" /></td>
</tr>
</tbody>
</table>

Answer: 

<table>
<thead>
<tr>
<th>Original Images</th>
<th>Drawings of the 1st Artist</th>
<th>Computer Generated Caricatures (in random order)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image12.png" alt="Image 12" /></td>
<td><img src="image12A.png" alt="Image 12A" /></td>
<td><img src="image12B.png" alt="Image 12B" /></td>
</tr>
<tr>
<td><img src="image12.png" alt="Image 12" /></td>
<td><img src="image12A.png" alt="Image 12A" /></td>
<td><img src="image12B.png" alt="Image 12B" /></td>
</tr>
</tbody>
</table>

Answer:
Question 2:

<table>
<thead>
<tr>
<th></th>
<th>Original Images</th>
<th>Drawings of the 2\textsuperscript{nd} Artist</th>
<th>Computer Generated Caricatures (in random order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>12</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Answer:

<table>
<thead>
<tr>
<th></th>
<th>Original Images</th>
<th>Drawings of the 2\textsuperscript{nd} Artist</th>
<th>Computer Generated Caricatures (in random order)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Answer:
Question 3:

<table>
<thead>
<tr>
<th>Original Images</th>
<th>Drawings of the 3rd Artist</th>
<th>Computer Generated Caricatures (in random order)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="11" alt="Image 1" /></td>
<td><img src="11" alt="Image 2" /></td>
<td><img src="11" alt="Image 3" /></td>
</tr>
<tr>
<td><img src="11" alt="Image 4" /></td>
<td><img src="11" alt="Image 5" /></td>
<td><img src="11" alt="Image 6" /></td>
</tr>
<tr>
<td><strong>11</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="12" alt="Image 1" /></td>
<td><img src="12" alt="Image 2" /></td>
<td><img src="12" alt="Image 3" /></td>
</tr>
<tr>
<td><img src="12" alt="Image 4" /></td>
<td><img src="12" alt="Image 5" /></td>
<td><img src="12" alt="Image 6" /></td>
</tr>
<tr>
<td><strong>12</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Answer:**
USE OF NEURAL NETWORKS IN AUTOMATIC CARICATURE GENERATION: AN APPROACH BASED ON DRAWING STYLE CAPTURE

Department of Computer Science, Loughborough University, UK
rupesh_shet@hotmail.com Fax: +44(0)1509 211586

Keywords: Caricature, Neural Networks, EDFM, Cascade Correlation, Pattern Recognition.

Abstract

Caricature is considered a rendering that emphasizes the distinctive features of a particular face. A formalization of this idea, Exaggerating the Difference from the Mean (EDFM) is widely accepted among caricaturists to be the driving factor behind caricature generation. However the caricatures created by different artists have distinctive features, which depend on their drawing style. No attempt has been taken in the past to identify these distinct drawing styles. Yet the proper identification of the drawing style of an artist will allow the accurate modelling of a personalised caricature. This work is the first attempt to use neural networks in this application area and have the potential to revolutionize existing automatic caricature generation technologies.

1 Introduction

Caricature is an art that conveys humour or sarcasm to people via drawing human faces. The basic concept is capturing the essence of a persons face by graphically exaggerating their distinctive facial features. Many approaches have been proposed in literature to automatically generate facial caricatures by computers [1-6]. Most of these approaches use fixed geometrical exaggerations based on simple image analysis techniques. Others use linguistic approaches, where exaggerations are based on variations linguistically requested by a user. However within the process of creating a caricature even a professional caricaturist would not be able to geometrically or linguistically quantify all the exaggerations he/she is likely to introduce. It is observed that these exaggerations often depend on the individual drawing style adopted by an artist. The fact that we are able to identify caricatures drawn by famous caricaturists, regardless of the original image, supports this observation. Unfortunately none of the existing state-of-the-art automatic caricature generation techniques attempt to capture the drawing style of an individual artist. Yet the accurate capture of this detail would allow more realistic caricatures to be generated. From the artists' point of view, it is difficult for them to explain how they draw caricatures. This is because the rules governing their drawing style are embedded in their subconscious mind and often unexplainable. Automatic identification of an artist's drawing style using state-of-the-art image analysis and artificial intelligence techniques could provide a solution for this.

The human brain has an innate ability of remembering and recognising thousands of faces it encounters during a lifetime, where most of the faces are metrically similar. Psychologists [5,6] suggested that human beings have a “mean face” recorded in their brain, which is an average of faces they encounter in life. A caricaturist compares one’s face with this so-called mean face, and uses their inborn talents to draw caricatures by exaggerating the distinctive facial features. This caricature drawing approach is widely accepted among psychologists and caricaturists [1,7]. Within the wider aspect of our research we are currently investigating the full automation of the above mentioned drawing style capture and related caricature generation process. The work presented in this paper limits the investigation to capturing the drawing style adopted by a caricaturist in exaggerating a single, selected facial component. It should be noted that capturing the drawing style adopted over a complete face is a challenging task due to the large number of possible variations and non-linearity of exaggerations that a caricaturist may adopt for different facial components. However non-linearity in exaggerations could be found even in the deformations made to a single facial component. This observation undermines previous research, which assumes semi-linear deformations over a single facial component such as an eye, mouth, chin, nose etc. Fortunately neural networks have the ability to capture the non-linear relationship between the input and output values in a training set. Within the research context of this paper we provide experimental results and analysis to prove that a Cascade Correlation Neural Network (CCNN) [11,12] can be trained to accurately capture the drawing style of a caricaturist in relation to an individual facial object. Further we use the results to justify that the trained CCNN could then be used to automatically generate a
caricature (drawn by the same artist) of the same facial component belonging to either the same original facial figure or of a different one.

For clarity of presentation the paper is organised as follows: section-2 introduces the CCNN and discusses its suitability for the application domain. Section-3 presents the proposed methodology for the use of CCNN in identifying the drawing style of an artist. Section-4 presents experimental results and a detailed analysis proving the validity of the proposed concepts in use in capturing the drawing style of a caricaturist. Finally section-5 concludes with an insight into further research that is currently being considered as a result of it.

2 The Cascade Correlation Neural Network

Artificial neural networks are the combination of artificial neurons that are similar to biological neurons [8,12,13]. These artificial neurons (simply called neurons here after) are usually connected in three layers. The first layer is an input layer, consisting of neurons that receive information (inputs) from the external environment. The second layer, which performs essential intermediate computations, is hidden from view (not directly visible from the external world) and is referred to as the hidden layer. The third layer is an output layer (target/output) that communicates the result of the weighted, summed output to the external environment or to the user. At the input layer, a linear input function computes the weighted sum of the inputs. Subsequently a non-linear transfer function transforms the weighted sum into final output values. Thus in general, all neural network architectures/topologies are based on the concept of input/output neurons, number of layers, a training function and transfer functions. Past research in neural network technology has resulted in the design of several architectures that are capable of solving specific problems. After testing and analysing various neural networks we found that the CCNN is the best for the application domain under consideration.

The CCNN [10-12] is a new architecture and is a generative, feed forward, supervised learning algorithm for artificial neural networks. It is similar to a traditional network in which the neuron is the most basic unit. However an untrained CCNN will remain in a blank state with no hidden units. Its output weights are trained until either the solution is found, or the progress stagnates. A hidden neuron is 'recruited' when training yields no appreciable reduction of error. Thus a pool of hidden neurons is created with a mixture of non-linear activation functions. The resulting network is trained until the error reduction halts. The hidden neuron with the greatest correspondence to the overall error is then installed in the network and the others are discarded. The new hidden neuron 'rattles' the network and significant error reduction is accomplished after each inclusion. Note that the weights of hidden neurons are static, i.e., once they are initially trained, they are not subsequently altered. The features they identify are permanently cast into the memory of the network, which means that it has the ability to detect the features from training samples. Preserving the orientation of hidden neurons allows cascade correlation to accumulate experience after its initial training session. Few neural network architectures allow this. The above features justify its use within the application domain under consideration. In addition the CCNN has several other advantages [11] namely: 1) It learns very quickly and is at least ten times faster than traditional back-propagation algorithms. 2) The network determines its own size and topology and retains the structure. 3) It is useful for incremental learning in which new information is added to the already trained network.

Once the architecture has been selected and the input signals have been prepared (unique properties have been found) the next step is to train the neural network. We use the Levenberg-marquardt backpropagation training function [12] due to its significant speed of operation. It would be unwise to design a network, train it and then put it into practise immediately. Its accuracy and capabilities should first be tested, evaluated and scrutinized. The testing process is known as validation. It can be said that the validation process is more important, as small errors could result in a misleading output from a network, which will be unreliable and incorrect. In section 4 we validate the use of the above network within the application domain under consideration.

3 Capturing the Drawing Style of a Caricaturist: The Proposed Methodology

Figure 1 illustrates the block diagram of the proposed drawing style capture algorithm. A facial component extractor module subdivides a given original facial image, its corresponding caricature drawn by the artist and the mean face into distinguishable components such as eye, nose, chin, mouth etc. Subsequently geometrical data from a given component of an original image and data from the corresponding component of the mean image are entered as inputs to the neural network module. The relevant data from the caricature component is entered to the module as the output. The above data is used to train the neural network. Once sufficient data points have been used in the above training process, we show that the neural network is able to predict the caricature of a novel image depicting the same facial component that was used in the training process. Below is a more detailed description of the processes involved.

Step 1: Generating Mean Face: For the purpose of our present research which is focused on a proof of concept, the mean face (and thus the facial components) was hand drawn for experimental use and analysis. However, in a real system one could use one of the many excellent mean face generator programs [15] made available in the World Wide Web.

Step 2: Facial Component Extraction/Separation: A simple image analysis tool based on edge detection, thresholding and thinning was developed to extract/separate various significant facial components such as, ears, eyes, nose and mouth from the original, mean and caricature facial
images (see figure 2). Many such algorithms and commercial software packages [14] exists that could identify facial components from images/sketches.

Figure 1: Proposed Drawing Style Capture Algorithm

Step 3: Creating Data Sets for Training the Neural network: Once the facial components have been extracted, the original, mean and caricature images of the component under consideration are overlapped, assuming an appropriate common centre point (see figure 3). E.g., for an eye, the centre of the iris could be considered the most appropriate centre point. Subsequently using cross sectional lines centred at the above point and placed at equal angular separations, the co-ordinate points at which the lines intersect the components are noted. This is done following a clockwise direction as noted by points 1, 2, ..., 8 of the caricature image data set of figure 3. Note that figure 3 is for illustration purposes only (not to scale) and thus may not represent an accurately scaled/proportioned diagram.

Figure 2: Facial component extraction from a caricature image

Figure 3: Creating Data Sets for Training
Step 4: Tabulating Data Sets: After acquiring the X-Y coordinate points as in step-3, they are tabulated as depicted in Table-1. The higher the number of cross sectional lines that are used, the more accurate the captured shape would be. However for clarity of presentation and ease of experimentation, we have only used four cross sectional lines in figure 3, which results in eight data sets.

<table>
<thead>
<tr>
<th>X1</th>
<th>Y1</th>
<th>X2</th>
<th>Y2</th>
<th>X3</th>
<th>Y3</th>
<th>X4</th>
<th>Y4</th>
<th>X5</th>
<th>Y5</th>
<th>X6</th>
<th>Y6</th>
<th>X7</th>
<th>Y7</th>
<th>X8</th>
<th>Y8</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>37</td>
<td>13</td>
<td>53</td>
<td>56.8</td>
<td>102</td>
<td>99</td>
<td>119</td>
<td>99</td>
<td>99</td>
<td>62</td>
<td>93</td>
<td>56.8</td>
<td>95</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>X1</td>
<td>Y1</td>
<td>X2</td>
<td>Y2</td>
<td>X3</td>
<td>Y3</td>
<td>X4</td>
<td>Y4</td>
<td>X5</td>
<td>Y5</td>
<td>X6</td>
<td>Y6</td>
<td>X7</td>
<td>Y7</td>
<td>X8</td>
<td>Y8</td>
</tr>
<tr>
<td>99</td>
<td>99</td>
<td>99</td>
<td>75</td>
<td>99</td>
<td>102</td>
<td>76</td>
<td>81</td>
<td>99</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>53</td>
<td>56.8</td>
<td>102</td>
<td>102</td>
<td>102</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Training Data Set

Step 5: Data Entry: Considering the fact that the neural network should be trained to automatically produce a caricature of a given facial component drawn by a particular artist, we consider the data points obtained from the caricature image above to be the output training dataset of the neural network. Furthermore the neural network is provided with the data sets obtained from the original and mean images to formulate input data. This follows the widely accepted strategy used by the human brain to analyse a given facial image in comparison to a known mean facial image.

Step 6: Setting up the Neural Network: We propose the use of the following training parameters for a simple, fast and efficient training process.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Network Name</td>
<td>Cascade Correlation</td>
</tr>
<tr>
<td>Training Function Name</td>
<td>Levenberg-marquardt</td>
</tr>
<tr>
<td>Performance Validation Function</td>
<td>Mean squared error</td>
</tr>
<tr>
<td>Number of Layers</td>
<td>2</td>
</tr>
<tr>
<td>Hidden Layer Transfer Function</td>
<td>Tan-sigmoid with one neuron at the start</td>
</tr>
<tr>
<td>Output Layer Transfer Function</td>
<td>Pure-linear with eight neurons</td>
</tr>
</tbody>
</table>

Table 2: Neural Network Specifications

Step 7: Testing: Once training has been successfully concluded as described above, the relevant facial component of a new original image is sampled and fed as input to the trained neural network along with the matching data from the corresponding mean component. In section-4 we provide experimental evidence in support of our proof of concept that a CCNN is able to capture the drawing style of a caricaturist.

4 Experiments and Analysis

Several experiments were designed and carried out to prove the suitability of using a CCNN to capture the drawing style of a caricaturist. The MATLAB neural network toolbox and associated functions [9] were used for the simulations. Three of these core experiments are presented and analysed in detail in this section. Note that experiments 1 and 2 use simple geometrical shapes for testing.

Experiment 1: In this experiment we train the CCNN using vertices of five equilateral triangles (numbered 1-5 in figure 4(a)). The first two vertices of each triangle, taken in a counter-clockwise direction are used as inputs to the neural network and the third vertex is used as the output. Subsequently given two ordered points in the Cartesian coordinate system, the neural network is used to predict the third point. It is found that the neural network is able to predict the third point to be a vertex of an equilateral triangle it forms with the two given points. In addition to this, the three points are arranged in the counter-clockwise direction, similar to the order used by the training datasets. The triangles numbered 6 and 7 in figure 4(a) illustrate the results of two such experiments. This proves that the CCNN is able to predict the orientation and direction with accuracy.

Figure 4: Experiment-1 Data (a) graphical (b) tabular
Experiment 2: The results of Experiment 1 proved that the CCNN is capable of accurately predicting orientation and direction. This experiment is designed to prove that it is able to accurately predict exaggeration in addition to the qualities tested under experiment 1. The four training objects denoted by 1-4 in figure 5(a) represent the training cases. In each training object, the innermost shape denotes the mean component, the middle shape denotes the original component and the outermost denotes the caricature component. Note that the exaggeration in one direction is much greater than in the other three directions for all training objects. Object 5 in figure 5(a) denotes the test case. The input shapes (mean and original) are illustrated by continuous lines and the output (i.e. generated caricature) shape is denoted by the dotted shape. Note that the CCNN has been able to accurately predict exaggeration along the proper direction, i.e. along the direction where exaggeration is the most when the original is compared with the mean in the test object.

Experiment 3: Experiments 1 and 2 were performed on basic shapes and proved that the CCNN is capable of accurately predicting orientation, direction and exaggeration. In this experiment we test CCNN on a more complicated shape depicting a mouth (encloses lower and upper lips). Figure 6 illustrates six training cases out of 20 cases used in the experiment. In each training case, the innermost shape corresponds to a mean mouth sampled at eight points. For all training and test cases the shape of the mean mouth has been maintained as constant [Note: To reduce experimental complexity, the sampling points were limited to 8. This does not undermine the experimental accuracy. However more sampling points would have allowed us to train the neural network on a more regular and realistic mouth shape.] The middle shape corresponds to the original mouth and the outermost shape represents the caricature mouth. Both these shapes have been sampled at 8 points as illustrated in training case 1 of figure 6. Note the non-linearity in exaggeration that is shown in the training cases across the shape of the mouth. Our set of 20 training cases was carefully selected so as to cover all possible exaggerations in all eight directions. This is a must in order for the CCNN to be able to predict exaggerations accurately in all of the eight directions.
Figures named “result 1-4” in figure 7 below, illustrate the test cases. They demonstrate that the successful training of the CCNN has resulted in its ability to accurately predict exaggeration of non-linear nature in all directions. Note that an increase in the amount of the training data set would result in an increase of the prediction accuracy for a new set of test data.

Figure 7: Testing CCNN on a real facial object under limited sampling – the test cases

4.2 Analysis: Use of CCNN in Automatic Caricature Generation

Our experiments above were designed to support the proof of concept that the CCNN can be used in capturing the drawing style of an artist and subsequent automatic caricature generation. Here we provide justifications as to why the experiments performed on limited shapes, with limited sampling would still prove enough evidence in support of the proposed idea.

Figure-8 illustrates the mean, original and caricature (drawn by two artists) images of a human eye. The original eye shows a noticeable difference in shape from the mean eye at the two ends. In the left end, the eye is curved up whereas at the right end it is curved down.

The drawing style of artist-1 shows no difference being made to the left side but a noticeable exaggeration to the difference (curved nature) in the right side. This could be a trait of this artist’s drawing style. I.e. the artist makes no exaggerations in any cartoon he draws, in the left corner of the eye, but exaggerates considerably in the right corner. The proposed CCNN based approach is able to learn this rule as proved by the results of experiments 2. Performing experiments on a larger set of original eyes (belonging to different people but caricatured by the same artist-1) will help improve prediction further. Using more sampling points around the surface of the eye (rather than 8 in our experiments) will increase the accuracy of approximating the actual shape of the eye.

In figure 8, the drawing style of artist-2 shows exaggerations being done at both ends of the eye. As justified above and supported by evidence from experiments 2 and 3, CCNN would be capable of accurately capture the drawing style of artist-2 as well. Given a new original eye, it would then be able to automatically generate the caricature, incorporating the artist’s style.

Figure 8: Comparison of the mean and an original human eye with a caricature eye drawn by two different artists

5 Conclusion

In this paper we have identified an important shortcoming of existing automatic caricature generation systems in that their inability to identify and act upon the unique drawing style of a given artist. We have proposed a Cascade-Correlation Neural Network based approach to identify the said drawing style of an artist by training the neural network on unique non-linear deformations made by an artist when producing caricature of individual facial objects. The trained neural network has been subsequently used successfully to generate the caricature of the facial component automatically. We have shown that the automatically generated caricature consists of various unique straits adopted by the artist in drawing free-hand caricatures.

The above research is a part of a more advanced research project that is looking at fully automatic, realistic, caricature generation of complete facial figures. One major challenge faced by this project includes, non-linearities and
unpredictabilities of deformations introduced in exaggerations done between different objects within the same facial figure, by the same artist. We are currently extending the work of this paper in combination with artificial intelligence technology to find an effective solution to the above problem.

References
NOVEL APPROACH TO NEURAL NETWORK BASED CARICATURE GENERATION


Department of Computer Science, Loughborough University, UK
K.H.Lai@lboro.ac.uk Fax: +44 (0)1509 635722

Keywords: Caricature Generation, Drawing Style, Artificial Intelligence, Neural network, Pattern Recognition.

Abstract

A caricature is defined as a funny drawing of someone that makes some of his/her distinct features appear exaggerated or more amusing. However, the caricatures of the same person created by different artists can be very different, since the artists' drawing styles play an important role [1]. Therefore, learning the drawing style of an artist provides the key to the computer-based automatic generation of professional caricature. Unfortunately, no caricature generation system in the past has attempted to address this issue with the aid of artificial intelligence technologies. In this paper, we propose an example-based caricature generation system with experimental results and detailed analysis to prove that neural networks can be used for capturing the drawing style of an artist. This work is the first system to use neural networks in generating caricature.

1 Introduction

Caricatures that provide humor and entertainment are common in our daily lives, with frequent appearances in magazines and newspapers (see figure 1).

Figure 1: Albert Einstein's caricature created by A. Hughes [1] with exaggerated hair, forehead and nose.

The caricaturing process is based on deforming the features of a face selectively. A caricaturist captures the essence of the subject, exaggerates the features and distorts the less important parts (or leaves them unchanged). These change the ratios among the subject's facial features and give a deeper impression to the viewers.

Unfortunately the talent of drawing caricature only exists in few people. This inborn talent, embedded in their subconscious mind, often makes it difficult for them to explain the craft of caricature generation to others. Therefore, caricature generation by computer has become a challenging research topic [2-7] especially as the underlying processes involved cannot be accurately explained. None of the existing computer-based approaches have attempted to learn the drawing style of an artist by using artificial intelligence technologies. Nevertheless, this is essential if caricatures of realistic expression are to be automatically generated.

On top of this, all of the existing state-of-the-art automatic caricature generation systems only provide linear exaggeration of facial components by scaling with a factor. However, non-linear exaggerations are an unavoidable key factor in professional caricature drawings. As illustrated in figure 2, a caricatured ear is not just exaggerated in size, but also non-linearly changed in shape. Inadequate handling of non-linear changes will lead to a decrease in reality. In our previous work [8], by using simple geometric shapes, we proved that a neural network can be used to capture the non-linear differences between two drawing objects. In this paper, we further extend this work and propose a novel algorithm to automatically generate photorealistic caricature.

Figure 2: An example of non-linear exaggeration. (a) Original image. (b) Corresponding caricature.

For clarity of presentation the paper is organised as follows: Section-2 presents the proposed methodology. Section-3 presents experimental results and detailed analysis proving the validity of the proposed algorithm's ability in capturing the drawing style of a particular caricaturist. Section-4 discusses the constraints and limitations of the project. Finally, section-5 concludes with an insight into further research that is currently being considered as a result of it.
2 Proposed methodology

The proposed drawing style capture algorithm is summarised as a block diagram in figure 3.

![Block Diagram](image)

Figure 3: Proposed drawing style capture algorithm.

2.1 Face selection and normalisation

We have used the AR face database [9] to provide facial images for testing. To maintain simplicity, only natural faces are chosen, any facial images with accessories such as glasses or hats have not been considered within the context of this work. Further, all chosen images are male, with short hair. Twelve facial images were chosen and two professional caricaturists were invited to draw caricatures. Each artist drew a set of caricatures that consisted of a drawing for each original image. An original image-caricature set drawn by one of our artists is shown in figure 4. In order to fully explore the drawing style of our artists, there were no drawing constraints to caricaturists, i.e., they were allowed to exaggerate or distort any facial component according to their styles, in contrary to previous methods. All original image-caricature pairs were subsequently normalised by using the distance between irises as the scaling factor. This is a widely accepted practice in face normalization in literature [2-7].

![Image of Original Images and Caricatures](image)

Figure 4: Twelve original images and their corresponding caricatures drawn by one of our artists.

2.2 Face modeling framework

To represent and control each facial feature, a geometric face model is required. The Facial Definition Parameters (FDPs) [10] of MPEG-4 standard has been chosen as the basis for the proposed work.

As the original design of FDPs in MPEG-4 is mainly used in facial animation, many feature points are designated for high motion facial components such as lips. However such an extensive amount of feature points are not necessary in facial caricaturing, so some of them are discarded for simplicity. Moreover, some subtle and static facial feature points, such as cheeks, are also removed. Yet, two new FDPs, 4.8 and 4.7 (see figure 5a), are added for better description of eyebrows. A comparison of the FDPs in MPEG-4 standard and our proposed FDPs with 46 feature points is presented in table 1. The proposed FDPs are also presented graphically in figure 5(a). Since the main focus of research is proving that the drawing style of an artist can be captured, facial feature points are marked manually on all original images and caricatures for improved accuracy and simplicity.
<table>
<thead>
<tr>
<th>Facial Components</th>
<th>Number of FDPs in MPEG-4</th>
<th>Number of Proposed FDPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Left Ear</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Right Ear</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Left Eyebrow</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Right Eyebrow</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Left Eye</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Right Eye</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Nose</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Mouth</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>74</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 1: Comparison of MPEG-4 FDPs and Proposed FDPs.

2.3 Mean face generation

Psychologists [11] suggest that everybody has a mental visualization of a "mean face", which is an average of the faces he/she has ever seen during lifetime. The driving factor of caricature drawing, widely agreed amongst the caricaturists is, exaggerating the difference from the mean (EDFM) face [12]. In this research, the original images 1 to 10 in figure 4 are used to generate a mean face by using the freeware, Morpher of [13]. The underneath technique is referred as morphing, which is not covered here [15]. The final generated mean face is illustrated in figure 5(b).

2.4 Relationships among original image, corresponding caricature and mean face

We hypothesize the caricature generation process as follows: When a caricaturist sees a face, he/she has the ability to identify the distinctive facial features by comparing it with the mean face hidden in his/her mind. The difference between an original face, \( O \), and a mean face, \( M \), is defined as \( \Delta S \) (The details of \( \Delta S \) calculation will be covered in section 2.5).

\[
\Delta S = O - M
\]  

(1)

Then based on the original image, the artist exaggerates the distinctive facial features intentionally to form a caricature of this particular face. Therefore, the difference between the original image, \( O \), and its corresponding caricature, \( C \), is the change made by the artist, which is defined as \( \Delta S' \) (The details of \( \Delta S' \) calculation will be covered in section 2.5).

\[
\Delta S' = C - O
\]  

(2)

The relationship between \( \Delta S \) and \( \Delta S' \) explains why the caricaturist makes this change, which refers to the drawing rules embedded in the artist's brain that governs his/her drawing style. The relationships among the original image, the corresponding caricature and the mean face are proposed in figure 6.

![Figure 5: (a) Proposed FDPs with 46 feature points. (b) Generated mean face from ten original images.](image)

2.5 Artificial neural network

Artificial neural network, commonly referred to as "neural network", is a type of artificial intelligence technologies that attempts to imitate the way a human brain works. There are over 10 billion biological neurons in a human brain, which cause us able to think, remember and learn. These human abilities can be simulated by connecting artificial neurons in a particular architecture to form a neural network [14].

The main reason of using neural network in this research is because of its effective learning ability. The network is capable to learn from the training set by constructing an input-output mapping for the problem automatically. Therefore, an understanding of how input is mapped to output is not necessary, which is perfectly fit for capturing the unexplainable relationship between \( \Delta S \) and \( \Delta S' \) that was discussed above. Moreover, a neural network has the ability to capture non-linear relationships from the training set and provides non-linear results [14]. This is suitable for capturing and mimicking non-linear exaggerations created by professional caricaturists.

The training set consists of both input and output values. To capture the relationship between \( \Delta S \) and \( \Delta S' \), the input and output to the neural network should be \( \Delta S \) and \( \Delta S' \) respectively. As illustrated in figure 7, an example of \( \Delta S \) and

![Figure 6: Relationship diagram of original image, corresponding caricature and mean face.](image)
A feature point is presented. Each oval represents the contour of an eye of a face, which is defined by eight feature points. They are normalized and overlapped with each other by using iris as the reference point.

Figure 7: An example of $\Delta S$ and $\Delta S'$ of a feature point.

$\Delta S$ of an original image-caricature pair can be calculated by subtracting the X and Y coordinates of all feature points of mean face ($X_M$ and $Y_M$) from their corresponding feature points of original image ($X_O$ and $Y_O$) respectively.

$$X_{AS} = X_O - X_M$$

$$Y_{AS} = Y_O - Y_M$$

Similarly, $\Delta S'$ of the same original image-caricature pair can be calculated by subtracting the X and Y coordinates of all feature points of original image ($X_O$ and $Y_O$) from their corresponding feature points of caricature ($X_C$ and $Y_C$) respectively.

$$X_{AS'} = X_C - X_O$$

$$Y_{AS'} = Y_C - Y_O$$

The above calculations are repeated on all remaining original image-caricature pairs. Subsequently, the X and Y training set tables are constructed. In this research, only the first ten original image-caricature pairs (1-10 of figure 4) are used for training. The remaining two original image-caricature pairs (11-12 of figure 4) are reserved for validations; further details will be discussed in section 3. In order to reduce the requirement of computer resource during the neural network training process, X and Y training sets are trained separately in two different neural networks with the same architecture and parameters.

2.6 Mesh warping

Mesh warping is the final module of the system that deforms an original test image into a caricature [15]. The module accepts an original facial image, a source face mesh and a target face mesh. These two meshes define how to deform an image into a caricature. In this project, the source face mesh is constructed by the XY coordinates of feature points of the original image, and the target face mesh is created by the XY coordinates of feature points generated from the neural networks (see section 2.5). Then the mesh warping module converts the original test image into a caricature, following the captured drawing style of the neural networks.

Table 2: Neural network architecture and training parameters.

<table>
<thead>
<tr>
<th>Architecture and Parameters</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of neural network required per artist</td>
<td>2</td>
</tr>
<tr>
<td>Neural Network Type</td>
<td>Feed-forward back-propagation</td>
</tr>
<tr>
<td>Training Function</td>
<td>Levenberg-Marquardt</td>
</tr>
<tr>
<td>Performance Validation Function</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>Performance Goal</td>
<td>0</td>
</tr>
<tr>
<td>Minimum Gradient</td>
<td>1e-010</td>
</tr>
<tr>
<td>Number of Layers</td>
<td>3</td>
</tr>
<tr>
<td>Hidden Layer Transfer Function</td>
<td>Tan-sigmoid</td>
</tr>
<tr>
<td>Output Layer Transfer Function</td>
<td>Pure-linear</td>
</tr>
<tr>
<td>Number of nodes in input layer</td>
<td>46</td>
</tr>
<tr>
<td>Number of nodes in hidden layer</td>
<td>46</td>
</tr>
<tr>
<td>Number of nodes in output layer</td>
<td>46</td>
</tr>
</tbody>
</table>

After training the neural networks, the next step is validations. (The validation results and full analysis will be covered in section 3.) Subsequently, any new original image can be sampled and fed as input to the trained neural networks for testing. The outputs will be the $X_{AS}$ and $Y_{AS}$ of the testing image, which can then be used to calculate the XY coordinates of the newly generated caricature.

3 Experimental results and analysis

In the first experiment, the neural networks are trained by using the drawings of our first artist. Afterward the $\Delta S$ of the remaining two original image-caricature pairs (11-12 of figure 4) pass through the trained neural networks separately for validations. The generated numerical results will then enter the mesh warping module for caricature deformations. The final photorealistic outputs are compared with the caricatures drawn by our first artist in figure 8 and 9 for validations.
Figure 8: (a) Original image II of figure 4. (b) Caricature of 8(a) drawn by our first artist. (c) Generated caricature of 8(a) from our system trained by the first artist’s drawings. In figure 8, by comparing the caricature drawn by our artist, 8(b), with the caricature generated by our system, 8(c), it can be shown that some of the drawing styles have been picked up successfully. First of all, the height of the forehead in 8(b) is shortened when compared with 8(a). A very similar forehead distortion happens in 8(c). Besides, similar exaggerations of the noses appear in both 8(b) and 8(c). Not only are the widths of noses increased, but also the lengths are elongated. Moreover, the changes of mouths in 8(b) and 8(c) are very close to each other. Both of them are stretched in width but not in height. Finally, the shapes of the eyes and eyebrows in 8(c) only slightly changed when compared with 8(a). These are similar to 8(b) as our artist did not modify these components vigorously, so these are also considered as the picked up drawing styles. Note that both hair and neck are not considered in our current system as they are not in the MPEG-4 FDPs standard, so these changes could not be captured.

Figure 9: (a) Original image 12 of figure 4. (b) Caricature of 9(a) drawn by our first artist. (c) Generated caricature of 9(a) from our system trained by the first artist’s drawings. In figure 9, by comparing the caricature drawn by our artist, 9(b), with the caricature generated by our system, 9(c), satisfactory results are demonstrated. Firstly, the ways of nose exaggeration in 9(b) and 9(c) are very close to each other, both of them are slightly bigger than the original image 9(a). On top of this, similar exaggerations of ears appear in both 9(b) and 9(c); their shapes and exaggeration ratios are almost the same. Furthermore, the sizes of mouths of 9(b) and 9(c) are both wider than 9(a) as the shapes remain unchanged. Moreover, the success of capturing drawing style is very obvious at the chins of 9(b) and 9(c), both of them are elongated heavily. Finally, both eyebrows and eyes of 9(c) are slightly wider than 9(a); these changes also match the drawing of our artist in 9(b).

Figure 10: (a) Original image 11 of figure 4. (b) Caricature of 10(a) drawn by our second artist. (c) Generated caricature of 10(a) from our system trained by the second artist’s drawings. In order to prove the validity of our proposed drawing style capture algorithm further, the second experiment has been carried out. The same methodology mentioned in section 2 with newly created neural networks has been applied on the drawings of our second artist. (Due to the page limitation, the training set caricatures drawn by our second artist are not shown here.) Figure 10 and 11 are the validations of the trained system. In figure 10(c), our caricature system captured the drawing style once again. The heavily elongated face and exaggerated nose are similar to the changes made by our second artist in 10(b). The slightly exaggerated mouth and elongated ears of 10(c) also match the drawings of 10(b). Likewise, in figure 11(c), the generated face is slightly elongated when compared with 11(a), this also happens in the caricature 11(b) that drawn by our second artist. Both size and shape changes of nose in 11(c) are very close to 11(b), the success of style capturing is obvious in this component. Finally, the slightly caricatured eyebrows, the widened eyes and remain unchanged mouth are similar in 11(b) and 11(c). Although a slightly tilted face and a dent at the chin cannot be seen in 11(c), a careful investigation revealed that these two drawing styles do not appear in the training set; hence our system cannot imitate them.

In the above analysis, justify the style similarity of two drawings could be subjective, and it is difficult to convince all the viewers that the styles of two caricatures are the same. After all, exactly capture and predict the style of a particular artist is an extremely difficult task. However, by comparing the results of the first and second experiments, the success of style capturing can be seen ultimately. When comparing computer generated caricatures 8(c) and 10(c) with the drawings of artists, 8(b) and 10(b), 8(c) looks closer to 8(b).
than 10(b). On the other hand, 10(c) looks similar to 10(b) rather than 8(b). Similarly, when comparing computer generated caricatures 9(c) and 11(c) with the drawings of artists, 9(b) and 11(b), 9(c) looks closer to 9(b) than 11(b). And 11(c) looks similar to 11(b) rather than 9(b). These comparisons demonstrated that the drawing rules of the first and second artists have been captured in the first and second experiments respectively, even though our system cannot predict and generate caricatures exactly the same as their drawings. The validation experiments and comparisons of system outputs for both artists provide satisfactory results. The generated caricatures, 8(c), 9(c), 10(c) and 11(c) illustrated that the trained neural networks have captured the drawing styles of our artists. As a result, the trained system can be put into practise.

4 Discussions

The proposed novel example-based neural network approach has captured the drawing style of a particular artist successfully, as proved in section 3. Training on a larger set of original image-caricature pairs (belonging to different people but caricatured by the same artist) will help improve prediction further. However, data collection has been a major obstacle of the project. It is unreasonable and time consuming to request an artist to draw too many caricatures for analysis, which also limits the applicability of the work. As a result, a key challenge of the project is to capture the drawing style of an artist from a limited dataset and still obtain satisfactory results.

Due to the fact that esthetic judgment of the resulting photorealistic caricatures could be a subjective task, a comprehensive subjective test is required to further analyse and evaluate more generated caricatures, which helps improve our system in the next stage. The above research is a part of a more advanced research project that is exploring the areas of caricaturing in different genders, ages and races. Besides, the hair and accessories will also be considered.

5 Conclusions

In this paper we have identified the drawbacks of existing caricature generation systems in their inability to capture non-linear relationship and the drawing style of a particular artist. We have proposed an example-based neural network approach to address the above issues by training neural networks on a small set of caricatures made by an artist. The proposed system has been evaluated by using the drawings of two professional caricaturists, both results proved that neural networks can be used for capturing the drawing style of an artist. The system is then capable to generate photorealistic caricatures with embedded styles automatically.

In the future, the trained neural networks can further be processed by artificial intelligence technology to extract linguistic rules from them [17]. This will help in the understanding and drawing style comparisons of different artists in both, artistic and psychological ways.

Acknowledgments

We would like to thank Mr. Yang and Mr. Abuhelga for their contributions in drawing caricatures for the project.

References

A FACIAL COMPONENT BASED HYBRID APPROACH TO CARICATURE GENERATION USING NEURAL NETWORKS

Department of Computer Science, Loughborough University, Leicestershire, LE11 3TU, UK.
K.H.Lai@lboro.ac.uk, E.A.Edirisinghe@lboro.ac.uk, P.W.H.Chung@lboro.ac.uk

ABSTRACT
A caricature is defined as a humorous drawing of a person that exaggerates or distorts certain distinctive features. However, the caricatures of the same person created by different artists can be very different, since the drawing styles of artists play an important role [1]. Therefore, learning the drawing style of an artist provides the key to the computer-based automatic generation of professional caricature. In our previous work [2-3], we proposed an example-based caricature generation system, which proved that neural networks can be used for capturing the drawing styles of caricaturists and successfully generating photo-realistic caricatures. Unfortunately, the quality of resulting caricatures was limited by the enormous demand of computational resources required by neural networks used; hence making the resolution of outputs a trade-off for computational power and memory limitations. In this paper, we propose a novel facial component based hybrid approach to resolve the above limitations and further improve the quality of the generated caricatures. Detailed experimental and subjective test results are provided and analyzed. This work is an extension of our previous system [3], which is the first attempt to use neural networks in generating caricatures.

KEY WORDS
Caricature generation, artificial intelligence, neural network, pattern recognition and drawing style

1. Introduction
Caricatures that provide humor and entertainment are common in our daily lives, with frequent appearances in magazines and newspapers. The caricaturing process is based on deforming the features of a face selectively. A caricaturist captures the essence of the subject, exaggerates the features and distorts the less important parts (or leaves them unchanged). These change the ratios among the subject's facial features and give a deeper impression to viewers. Unfortunately, the talent of drawing caricature only exists in few people. This inborn talent, embedded in their subconscious mind, often makes it difficult for them to explain the craft of caricature generation to others. Therefore, caricature generation by a computer has become a challenging research topic [4-8] especially as the underlying processes involved cannot be accurately explained. However, none of these state-of-the-art computer-based systems has attempted to learn the drawing style of an artist by using artificial intelligence technologies. On top of this, all of the existing automatic caricature generation systems only provide linear exaggeration of facial components by scaling with a factor. Yet, in a professional caricature drawing, an artist does not only exaggerate the sizes of facial components, but also changes their shapes non-linearly. Hence, handling of non-linear changes cannot be ignored. In view of the above factors, by using simple geometric shapes, we demonstrated that a neural network can be used to capture and predict the non-linear differences between two drawing objects [2]. We further extended the work and proposed an example-based caricature generation system, which proved that neural networks can be used for capturing the drawing styles of artists and generating photo-realistic caricatures with their styles embedded [3]. Nonetheless, the quality of caricatures generated by our system is constrained by the high demand for computer resources. When the resolutions of output caricatures increase, the computational power and memory required by neural networks increase exponentially [14]. As a result, this limitation leads to a decrease of reality in generated caricatures. In this paper, we further extend our system [3] and propose a novel facial component based hybrid approach to automatically generate improved quality photorealistic caricatures, without increasing the requirement of computational resources. For clarity of presentation the paper is organised as follows: Section-2 presents the proposed methodology. Section-3 presents experimental and subjective test results with detailed analysis to prove that the proposed algorithm has improved the quality of our system. A full comparison of generated caricatures with our previous work [3] is provided. Section-4 discusses and concludes with an insight into further research that is currently being considered as a result of it.
2. Proposed methodology

The proposed component based drawing style capture algorithm is summarized as a block diagram in figure 1.

![Block diagram of the proposed component based drawing style capture algorithm](image)

Figure 1: Proposed Component Based Drawing Style Capture Algorithm.

2.1 Face Selection and Normalization

We have used the AR face database [9] to provide facial images for testing. To maintain simplicity, only natural faces are chosen, any facial images with accessories such as glasses or hats have not been considered within the context of this work. Further, all chosen images are male, with short hair. Twelve facial images were chosen and several professional caricaturists were invited to draw caricatures, of which caricatures from two artists are used for our theoretical presentation and experimentation in this paper. Each artist was invited to draw a set of caricatures for each of the twelve original images. An original image-caricature set drawn by one of our artists is shown in figure 2. In order to fully explore the drawing styles of our artists, there were no drawing constraints to them, i.e., they were allowed to exaggerate or distort any facial component according to their styles. All original image-caricature pairs were subsequently normalized by using the distance between irises as the scaling factor. In literature, this is a widely accepted practice in face normalization [4-8].

![Twelve original images and their corresponding caricatures](image)

Figure 2: Twelve Original Images and Their Corresponding Caricatures Drawn by One of Our Artists.

2.2 Face Modeling Framework

To represent and control each facial feature, a geometric face model is required. In our previous caricature generation system [3], a simple face model with 46 facial feature points was proposed, which was defined according to the Facial Definition Parameters (FDPs) [10] of MPEG-4 standard. By using such a simple model to represent the variety present in human faces is obviously not enough. However, the design of this simple face model was for saving computer resources and increasing simplicity. In this paper, a new geometric face model with 145 FDPs is proposed for a better description of each facial component, which is constructed based on the simple face model. As the number of FDPs increases, the resolution of representing an image increases. A comparison of FDPs of the simple face model and the proposed model is presented in table 1. The proposed FDPs are also presented graphically in figure 3(a). The main focus of our present research is proving that without increasing the computational resources, the proposed algorithm can still improve the quality of generated caricatures. Thus, all facial feature points are marked manually for increased accuracy and simplicity.

<table>
<thead>
<tr>
<th>Facial Components</th>
<th>No. of FDPs in Simple Model</th>
<th>No. of FDPs in Proposed Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>Both Ears</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>Both Eyebrows</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>Both Eyes</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>Nose</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Mouth</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
<td>145</td>
</tr>
</tbody>
</table>

Table 1: Comparison of FDPs of the Simple Face Model [3] and the Proposed Face Model.

![Proposed FDPs with 145 feature points](image)

(a)

![Generated mean face from ten original images](image)

(b)

Figure 3: (a) Proposed FDPs with 145 Feature Points. (b) Generated Mean Face from Ten Original Images.
2.3 Mean Face Generation

Psychologists [11] suggest that everybody has a mental visualization of a "mean face", which is an average of the faces he/she has ever seen during lifetime. The driving factor of caricature drawing, widely agreed amongst the caricaturists is, exaggerating the difference from the mean (EDFM) face [12]. In this research, the original images 1 to 10 in figure 2 are used to generate a mean face by using the freeware, Morpher of [13]. The underlying technique is referred to as morphing, which is not covered here [15]. The final generated mean face is illustrated in figure 3(b).

2.4 Relationships among Original Image, Corresponding Caricature and Mean Face

We hypothesize the caricature generation process as follows: When a caricaturist sees a face, he/she has the ability to identify the distinctive facial features by comparing it with the mean face hidden in his/her mind. The difference between an original face, \( O \), and a mean face, \( M \), is defined as \( \Delta S \) (The details of \( \Delta S \) calculation will be covered in section 2.5).

\[
\Delta S = O - M
\]  
(1)

Subsequently based on the original image, the artist exaggerates the distinctive facial features intentionally to form a caricature of this particular face. Therefore, the difference between the original image, \( O \), and its corresponding caricature, \( C \), is the change made by the artist, which is defined as \( \Delta S' \) (The details of \( \Delta S' \) calculation will be covered in section 2.5).

\[
\Delta S' = C - O
\]  
(2)

The relationship between \( \Delta S \) and \( \Delta S' \) explains why the caricaturist makes this change, which refers to the drawing rules embedded in the artist's brain that govern his/her drawing style. The relationships among the original image, the corresponding caricature and the mean face are proposed in figure 4.

![Figure 4: Relationship Diagram of Original Image, Corresponding Caricature and Mean Face.](image)

By capturing the relationships between \( \Delta S \) and \( \Delta S' \) from original image-caricature pairs (belongs to a particular artist), the drawing style of the artist can be learnt, the details will be covered in section 2.5. However, the relationship is always non-linear and difficult to describe in written language precisely. As a result, we adopt a neural network to accomplish this task.

2.5 Artificial Neural Network

Artificial neural network is a type of artificial intelligence technologies that attempts to imitate the way a human brain works, which can simulate human abilities such as thinking, remembering and learning by connecting artificial neurons in a particular architecture [14]. The main reason for using a neural network in this research is its effective learning ability. The network is capable to learn from the training set by constructing an input-output mapping for the problem automatically. Therefore, an understanding of how the input is mapped to the output is not necessary, which is perfectly fit for capturing the unexplainable relationship between \( \Delta S \) and \( \Delta S' \) as discussed above. Moreover, a neural network has the ability to capture non-linear relationships from the training set and provides non-linear results [14]. This is suitable for capturing and mimicking non-linear exaggerations created by professional caricaturists.

The training set consists of both input and output values. To capture the relationship between \( \Delta S \) and \( \Delta S' \), the input and output to the neural network should be \( \Delta S \) and \( \Delta S' \) respectively. As illustrated in figure 5, an example of \( \Delta S \) and \( \Delta S' \) of a facial feature point is presented. Each oval represents the contour of an eye of a face, which is defined by eight facial feature points. They are normalized and overlapped with each other by using the iris as the reference point.

![Figure 5: Example of \( \Delta S \) and \( \Delta S' \) of a Feature Point.](image)

\( \Delta S \) of a facial component of an original image-caricature pair can be calculated by subtracting the X and Y coordinates of feature points of mean face that belong to the component \( (X_M, Y_M) \) from their corresponding feature points of original image \( (X_O, Y_O) \) respectively.

\[
X_{\Delta S} = X_O - X_M \\
Y_{\Delta S} = Y_O - Y_M
\]  
(3) and (4)

Similarly, \( \Delta S' \) of the facial component of the same original image-caricature pair can be calculated by...
subtracting the X and Y coordinates of feature points of original image that belong to the component \((X_0\) and \(Y_0\)) from their corresponding feature points of caricature \((X_c\) and \(Y_c\)) respectively.

\[
X_{st} = X_c - X_0 \\
Y_{st} = Y_c - Y_0
\]

(5) \hspace{1cm} (6)

The above calculations are repeated on all remaining original image-caricature pairs. Subsequently, the X and Y training set tables of a component are constructed. The whole process is then repeated until training set tables of all components are calculated. In this research, only the first ten original image-caricature pairs (1-10 of figure 2) are used for training. The remaining two original image-caricature pairs (11-12 of figure 2) are reserved for validations; further details will be discussed in section 3.

In order to reduce the requirement of computer resources during the neural network training processes, X and Y training sets are trained separately in two different neural networks with the same architecture and parameters. Training the same type of neural networks in different ways can have different results. In our previous work [3], all the facial feature points of the face were trained at the same time. Therefore, the numbers of neurons in both input and output layers are the same as the total number of FDPs in the simple face model of table 1, which is 46 points. We term this algorithm the “Whole Face Based Training” (WFBT) approach. However, in the proposed face model of table 1, the number of FDPs is about three times the simple face model, the training time and the requirement of computational resources increase exponentially [14]. This is beyond the capability of a desktop computer (Intel P4 2.8GHz CPU with 1GB RAM). In view of this, we propose a novel “Facial Component Based Training” (FCBT) approach to address this issue. Each facial component is trained in an individual neural network; this not only decreases the computational resource requirement as the number of feature points per training decreases, but also increases the resolution of each component because the number of feature points used to represent a component increases (see table 1). As a result, with lower computer resources, higher quality of resulting caricatures is expected after combining the generated components together.

Once the training sets have been prepared, the next step is to define the neural networks. In this project, all experiments were carried out by using the MATLAB neural network toolbox [16]. We used feed-forward back-propagation network with only one hidden layer as the architecture. The numbers of nodes of input, hidden and output layers of a neural network were the same as the number of FDPs that proposed to define the particular facial component in table 1. The neural networks were trained by using the Levenberg-Marquardt [16] algorithm without momentum. The mean squared error was used as the performance function and the performance goals were set to zero. The trainings stopped once the performances were minimized to the goals or the gradients of performances were less than the minimum gradient parameter, which means the error slope is close to zero and further training cannot reduce the error by much [16]. A summary of the neural network architecture and the training parameters are shown in table 2.

<table>
<thead>
<tr>
<th>Architecture and Parameters</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Neural Networks</td>
<td>2</td>
</tr>
<tr>
<td>Required per Facial Component</td>
<td>12</td>
</tr>
<tr>
<td>Number of Neural Networks</td>
<td>Feed-forward</td>
</tr>
<tr>
<td>Required per Artist</td>
<td>Back-propagation</td>
</tr>
<tr>
<td>Neural Network Type</td>
<td>Levenberg-Marquardt</td>
</tr>
<tr>
<td>Performance Function</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>Performance Goal</td>
<td>0</td>
</tr>
<tr>
<td>Maximum Gradient</td>
<td>1e-010</td>
</tr>
<tr>
<td>Maximum Number of Epochs</td>
<td>1000</td>
</tr>
<tr>
<td>Number of Layers</td>
<td>3</td>
</tr>
<tr>
<td>Hidden Layer Transfer Function</td>
<td>Tan-sigmoid</td>
</tr>
<tr>
<td>Output Layer Transfer Function</td>
<td>Pure-linear</td>
</tr>
<tr>
<td>Number of Neurons in Input, Hidden and Output Layers</td>
<td>Number of FDPs of the Facial Component</td>
</tr>
</tbody>
</table>

Table 2: Neural Network Architecture and Parameters.

After training the neural networks, the next step is validations. The validation results and a full analysis will be covered in section 3. Subsequently, \(\Delta S\) of any new original facial component can be sampled and fed as input to its corresponding trained neural networks for testing. The outputs will be the \(X_{st}\) and \(Y_{st}\) of the testing component, which can then be used to calculate the XY coordinates of the newly generated component.

### 2.6 Facial Components Positioning

In our previous work [3], the system was trained by the WFBT approach as discussed in 2.5. The advantage of training all the facial components together is that neural networks are capable to learn the exaggerations of components and also their relative positions, the facial components in the generated caricature are always positioned properly. In this paper, the proposed FCBT approach trains the facial components separately. Although a higher quality of facial component shape predictions is expected, the information of relative positions of components is lost. Therefore, overlapping facial components exists in the resulting caricature if the generated components are combined together without preprocessing. This commonly happens as a heavily elongated nose overlaps with an unchanged mouth, or exaggerated eyes overlaps with unchanged eyebrows since they are very close to each other. In order to remedy this problem, we propose a hybrid algorithm that uses the caricature generated from the WFBT approach as a facial component position template. Subsequently all the components generated from the FCBT approach are shifted to the corresponding positions according to the template. This FCBT hybrid approach provides the final generated caricature with high quality and proper positioned facial components. Further details and comparisons of results from two different approaches will be covered in section 3.
2.7 Mesh Warping

Mesh warping is the final module of the system that deforms an original test image into a caricature [15]. The module accepts an original facial image, a source face mesh and a target face mesh. These two meshes define how to deform an image into a caricature. In this project, the source face mesh is constructed by the XY coordinates of FOPs of the original image, and the target face mesh is created by the XY coordinates of FOPs generated from the neural networks of different facial components (see section 2.5) after positioning (see section 2.6). Then the mesh warping module converts the original test image into a caricature, following the captured drawing style of the neural networks.

3. Experimental Results and Analysis

In the first experiment, neural networks are trained by using the facial components of our first artist’s drawings (1-10 of figure 2). Afterward the ΔS of facial components of the remaining two original image-caricature pairs (11-12 of figure 2) pass through their corresponding trained neural networks for validations. The generated numerical results of different facial components are positioned together by using the proposed hybrid algorithm, and then enter the mesh warping module for caricature deformations. The final photorealistic outputs are compared with the caricatures drawn by our first artist and also the caricatures generated by the WFBT approach in figure 6 and 7 for validations. Note that the caricatures generated by the WFBT approach, 6(c), 7(c), 8(c) and 9(c), have captured and embedded the drawing styles of our artists, which have already been proved in [3].

In figure 6, by comparing the caricature drawn by our artist, i.e. 6(b), with the caricatures generated by the WFBT approach, 6(c), and the FCBT hybrid approach, 6(d), most facial components of 6(b) are closer to those of 6(d) than to those of 6(c). First of all, the shape of nose of 6(b) is more similar to that of 6(d) than to that of 6(c), as specifically shown by the lower part. Besides, the shapes of ears of 6(b) are closer to the regularly exaggerated ears of 6(d) than to that of 6(c). Furthermore, the sizes and shapes of the mouth, eyes and eyebrows of 6(b) are slightly closer to 6(d) than to that of 6(c). Finally, the face shape of 6(b) looks similar to 6(d) than to that of 6(c).

In figure 7, by comparing the caricature drawn by our artist, i.e. 7(b), with the caricatures generated by the WFBT approach, 7(c), and the FCBT hybrid approach, 7(d), a more satisfactory result is illustrated in 7(d). Firstly, the sizes of both eyes in 7(b) are closer to the more exaggerated eyes of 7(d) than to that of 7(c). Moreover, the shape of the face in 7(d) is better than 7(c) as the sizes of forehead and jaws are more similar to that of 7(b). Besides, the way of eyebrow distortions in 7(b) is nearer to that of 7(d) than to that of 7(c). Finally, the shapes of ears and nose in 7(b) are also slightly closer to that of 7(d) than to that of 7(c).

In the second experiment, the same methodology proposed in section 2 was applied on the drawings of the second artist. (Due to the page limitation, the training set caricatures drawn by our second artist are not shown here.) Figure 8 and 9 are the validations of the trained system. Figure 8 further illustrates that the FCBT hybrid approach performs better than the WFBT approach. Firstly, the shapes of both ears of 8(b) are closer to that of 8(d) than to that of 8(c), as the widths of ears show clearly. Also, the degree of exaggeration of the face in 8(d) is more
appropriate than 8(c), when compared with 8(b). In addition, the shapes of both eyebrows and eyes in 8(b) are more similar to those of 8(d) than to those of 8(c). Likewise, in figure 9(d), the ears are elongated in a balanced way, which match the ears in 9(b) but not 9(c). Moreover, the sizes of eyes in 9(b) are closer to those of 9(d) than to those of 9(c). Finally, the shape of the face in 9(b) is much similar to 9(d) than to 9(c), this is demonstrated on both jaws visibly. Although a slightly tilted face and a dent at the chin of 9(b) still cannot be seen in 9(d), a careful investigation revealed that these two drawing styles do not appear in the training set; hence our system has not been able to imitate them.

In the above validation experiments it can be concluded that the FCBT hybrid approach provides more satisfactory results than the WFBT approach. However, aesthetic analysis and justification of the resulting photorealistic caricatures could be subjective, and it may also be difficult to convince all the viewers that the FCBT hybrid approach performs better than the WFBT approach. As a result, a subjective test was carried out to further confirm the success of the proposed FCBT hybrid approach. A questionnaire for comparing the results of the two different approaches, with the caricature drawings of the artists was setup. Thirty naïve volunteers, with respect to the experimental hypothesis being tested, were invited to participate. For the first artist, 77% and 70% participants expressed that the caricatures drawn by the artist, 6(b) and 7(b), are more similar to the caricatures generated by the FCBT hybrid approach, 6(d) and 7(d), than the WFBT approach, 6(c) and 7(c), respectively. Similarly, 80% and 77% participants answered that the caricatures drawn by our second artist, 8(b) and 9(b), are closer to the caricatures generated by the FCBT hybrid approach, 8(d) and 9(d), than the WFBT approach, 8(c) and 9(c), respectively. In conclusion average 76% participants stated that caricatures generated by the FCBT hybrid approach are closer to the drawings of our artists than the closeness shown by caricatures generated by WFBT approach to the drawings of our artists. Therefore it can be concluded that the FCBT hybrid approach performs better than the WFBT approach in caricature generation.

4. Discussions and Conclusions

Training on a larger set of original image-caricature pairs (belonging to different people but caricatured by the same artist) will help improve prediction further. However, data collection has been a major obstacle of the project. It is unreasonable and time consuming to request an artist to draw too many caricatures for analysis, which also limits the applicability of the work. As a result, the key challenges of the project are to capture the drawing style of an artist from a limited dataset, as done in [3], and also improve the quality of outputs as much as possible, even the computer resources are constrained.

In this paper we have identified the limitation of our previous work, the “Whole Face Based Training” approach caricature generation system that proposed in [3]. We have proposed a novel “Facial Component based Training” hybrid approach to resolve the limitation of inadequate computational resources and also improved the quality of generated caricatures. The system has been evaluated by the drawings of two professional caricaturists; both results proved that under the same computational resource condition, the FCBT hybrid approach performs better than the WFBT approach, as the generated caricatures are closer to the artists’ drawings. This has been further confirmed by subjective test results. In the future, the above research can be extended to explore caricatures with different genders, ages and races.

Acknowledgements

We would like to thank Mr. Yang and Mr. Abuhelga for their contributions in drawing caricatures for the project.

References

Automatic Photorealistic Caricature Generation: Two Approaches Based on Neural Networks

Ka H. Lai, Paul W. H. Chung and Eran A. Edirisinghe

Abstract—A caricature is defined as a humorous drawing of a human facial figure that makes some of its distinct features appear exaggerated. It is easily observed that the exaggerations made by different artists on facial components are often different and are non-linear. This uniqueness of the exaggerations signifies the drawing style of an artist, but has unfortunately been ignored in the design of existing computer based automatic caricature generation systems. Nevertheless, learning the unique drawing style and modeling the non-linear exaggerations distinct to an artist provide the key to the computer based automatic generation of professional caricature. In this paper, we propose two example-based caricature generation systems where a neural network is used to capture the drawing style of an artist and a morphing tool is subsequently used to automatically create new caricatures. We provide experimental results and detailed analysis to prove that our approach is capable of accurately capturing the drawing style of an artist and is able to thus create photorealistic caricature.

Index Terms—Caricature, Drawing Style, Mean Face, Neural Network, Mesh Warping, Pattern Recognition

I. INTRODUCTION

CARICATURE has been a common and popular artistic approach that has been used to convey humor, especially in magazines and newspapers. The process of caricaturing is based on selectively deforming the facial features. A professional caricaturist is able to mentally capture the essence of the unique features of a subject (i.e., of the facial image) and exaggerate them [1]. Less unique features are often left unchanged. These artistic alterations change the relative ratios of the subject’s facial feature and give a deeper impression to viewers. Hence, the information in the caricature is often considered richer than that of the original image. This has been verified by psychological experiments that have proved that the time required for recognizing a caricatured image is less than the corresponding veridical ones [2],[3].

Though the ability to recognize a person in a caricature widely exists amongst humans, the ability of drawing caricatures only exists in a few people, and has thus been considered as an innate talent. Further, several studies have revealed that it is difficult for a caricaturist to explain the process of drawing a caricature as the associated drawing rules are embedded in their minds and often appears to be fuzzy. This provides numerous challenges to the researchers involved in automatic, computer based generation of caricature. Susan E. Brennan [4], was the first to attempted computer-assisted 2D caricature generation. She adopted the concept of "mean face", well recognized by both caricaturists to be instrumental in drawing caricatures [5] and psychologists in determining facial uniqueness and attractiveness [6]. In psychology the mean face refers to the average of faces one comes across during one’s lifetime. In the approach proposed in [4] the author first calculated a mean face by averaging ten males, and subsequently exaggerated the facial difference of a target subject away from the mean by scaling with a selected factor. The resulting image was presented as the computer generated caricature of the subject. Based on this idea, many approaches have been proposed to generate facial caricatures by computers, automatically [7],[8],[9],[10],[11],[12],[13]. Most of them are geometric approaches that are based on analysing the facial feature points of a face. A few of them are linguistic approaches, which apply fuzzy logic technologies. L. Liang et al [14] attempted to capture the drawing style of an artist by using a prototype-approach based on a large dataset. However, the generated caricatures are very limited as only one facial component can be exaggerated in each of the results. In this paper, the Picasso system proposed in [8] is used as a benchmark to compare with caricatures generated by our approaches (see section III), due to its popularity and easy implementation.
Unfortunately, none of the existing computer based caricature generation approaches are able to capture the often unique drawing style of an artist. Nevertheless capturing the drawing style of an artist helps in the understanding of the mystery caricature, signifying the unique drawing style of the artist. Such an approach will find applications in situations where a caricaturist will require the help of a computer based system to generate artwork unique to their style with the intention of reducing their workload or in case of disability. Further such a system will help society to maintain the artwork of a famous caricaturist, after his/her unfortunate demise.

Existing state-of-the-art automatic caricature generation systems only provide linear exaggerations of facial components by scaling the whole component by a constant factor. However our careful observations revealed that, non-linear exaggeration even within a single component is a key factor in professional caricature drawing. Figure 2 illustrates that a caricatured ear is not just exaggerated in size, but also non-linearly changed in shape and orientation. Therefore inadequacy of handling non-linear exaggerations will lead to a decrease in realism of the generated caricature. In [13] we conceptually proved that a neural network is capable of capturing non-linear exaggerations introduced into a number of basic geometrical shapes. In this paper, we further extend this work by using a neural network to capture the unique drawing style of a caricaturist. Subsequently we use the captured drawing style to create realistic caricature from new target images. We propose two approaches: one based on exaggerations made to the entire face and the other based on independent exaggerations to individual components of the face. We provide experimental results for both approaches based on caricatures drawn by three different artists. The results are compared to that of an existing approach and conclusions are drawn on the effectiveness of the proposed approaches.

For clarity of presentation, the paper is organised as follows: section-2 presents the proposed methodologies for the identification of the drawing style of an artist and subsequent caricature creation. Section-3 presents experimental results and a detailed analysis proving the validity of the proposed concepts used in capturing the drawing style. Section-4 discusses the practical challenges faced by methodologies adopted, methods used to overcome them and possible application domains of the proposed work. Finally section-5 concludes with an insight into further research that is currently being considered as a result of the research presented in this paper.

II. PROPOSED METHODOLOGIES

Two approaches to automatic caricature generation are proposed. The first approach considers the entire face as a single object. Therefore selected feature points corresponds to facial features and the relative positioning of the facial components such as eyes, nose and mouth, are maintained. In the second approach, the first approach is used on individual facial components rather than on the entire face, i.e. feature points now corresponds to the features of individual components and the exaggerations made by a caricaturist to each facial component is compared against the mean of the shape of the corresponding facial component across the set of test facial images. As this approach does not maintain the relative positioning of the facial components on the face, a final facial component positioning stage is used for the creation of the computer generated caricature. Due to the striking similarity of the fundamental ideas underpinning the two approaches (entire face vs. individual facial component) we first present the entire face-based approach and then discuss the additional design considerations of the component-based approach.

![Fig. 2. An example of non-linear exaggeration in caricature drawing. (a) Original image. (b) Corresponding caricature.](image-url)

![Fig. 3. Proposed entire face-based approach.](image-url)
II.1 Entire Face-Based Approach

A. Overview

A block diagram of the proposed entire face-based automatic caricature generation algorithm is illustrated in figure 3. Firstly, a number of data capture/preparation stages are required for each facial image, such as face selection, caricature drawing, normalization and manual marking of feature points. Secondly, after capturing and appropriate pre-processing, the data associated with all facial images, a mean face generation module is used for calculating the mean face. Subsequently, within the training phase, the geometrical differences between the original images and the mean face (calculated based on the feature points) are fed to the neural network as inputs; while the geometrical differences between the corresponding caricatures and the original images will be considered the outputs of the neural network. Once the training phase is completed, the geometrical differences between a test image and the mean face are fed to the neural network for testing. Finally, the corresponding output after converting to \((x,y)\) coordinates is forwarded to a mesh warping module that deforms the original test image into a caricature. In the following sections we provide more detailed descriptions of the processes introduced above.

B. Face Selection, Caricature Drawing and Normalisation

The AR face database [16] was chosen for the project to provide the original facial images. This database consists of set of high resolution facial images, taken under controlled conditions and has hence been used as a popular data set in previous research. Further the fixed pose maintained in capturing the facial images of the database, greatly decreases the complexity of normalisation. We have further limited our data set to a collection of male facial images with short hair and no accessories, such as ‘spectacles’. Twelve images from the AR face database were selected in total, and three professional caricaturists were invited to produce caricature drawings. In order to fully explore the drawing styles of our artists, there
were no drawing constraints to them, i.e., they were allowed to exaggerate or distort any facial component according to their styles.

Although the original images and the caricatures are captured under similar pose and scale, they are normalized to further convert them into same scale and inclination level, so that the resulting images will be accurately comparable with each other. The normalisation approach proposed by Susan E. Brennan [4], which is widely used in previous research in automatic caricature generation and psychology [3],[7],[8],[9],[10],[11],[12],[13] was adopted. It assumes that the distance between two irises is a constant for all people. The advantages of choosing iris separation as the scaling factor are that it can be measured accurately and is independent of facial expressions as compared to other possible facial measures. Once the XY coordinates of both irises are recorded, the distance between them, \( d \), is given by

\[
d = \sqrt{(X_l - X_r)^2 + (Y_l - Y_r)^2}
\]  

(1)

where \( l \) and \( r \) corresponds to the left and right irises respectively. Moreover, the locations of irises can be used to determine the inclination level of a face, \( \Theta \), and function as the rotation factor. The formula is given by

\[
\Theta = \tan^{-1}\left(\frac{Y_l - Y_r}{X_l - X_r}\right)
\]  

(2)

The scaling factor is used to scale the face as the rotation factor is used to rotate the face. The final normalised original image-caricature pairs for the three artists are illustrated in figure 4.

C. Selecting Feature Points

To represent and control each facial feature, a geometric face model is required. A number of face models have been proposed in literature [4],[17],[18],[19] during the last three decades. However, since the standardization of MPEG-4 [19],[20],[21] in 1998, this authoritative standard has provided the de-facto methodologies in face modelling. Within the context of our present research the Facial Definition Parameters (FDPs) [19] of MPEG-4 standard are used as the basis for feature point selection and facial representation. Unfortunately according to the FDPs of MPEG-4, only three feature points are defined for the hair, which are obviously not sufficient for accurately representing most hairstyles. Under this limitation, the accurate modeling of hair within the context of current research becomes a tedious task. Therefore for simplicity, the caricaturing of the hair has been ignored within our present design. Further depending on the need for added representation accuracy of more significant facial components, additional feature points are added, for example points 4.7 and 4.8 (see figure 5(b)) are added to the right and left eyebrows, respectively. On the contrary, the design of FDPs in MPEG-4 standard is mainly focused on facial animation. Therefore a large number of feature points are designated to accurately represent high motion facial components such as the mouth and lips. Such representation accuracy is not necessary in an automatic caricaturing system. Further within our present research context, some subtle and static facial feature points, such as those representing hairlines and the cheeks, are excluded from consideration. As a result, the required computer resources during the neural network training process (see subsection F) can be maintained at lower level. In table I, a summary of the number of FDPs defined in MPEG-4 standard for each facial component is compared against the number of FDPs used in the proposed. The original MPEG-4 FDPs and the proposed FDPs are illustrated in figure 5. Note that for clarity of representation some facial feature points are left unnumbered. Further note that feature points such as 5.2, 5.4 (of the cheek) and 11.4 (hairline) in figure 5(a) has been ignored in 5(b).

<table>
<thead>
<tr>
<th>Region</th>
<th>MPEG-4 FDPs</th>
<th>Proposed FDPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Left Ear</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Right Ear</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Left Eyebrow</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Right Eyebrow</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Left Eye</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Right Eye</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Nose</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Mouth</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>74</td>
<td>46</td>
</tr>
</tbody>
</table>

A manual marking strategy is adopted for the selection of FDPs of the normalized original and caricature images. Our preliminary investigations have revealed that under the context of our present research, which focuses on proving that the proposed approach is capable of capturing the drawing style of an artist and is then able to automatically generate caricatures of the artist, this is a decision that provides reasonable accuracy.

<table>
<thead>
<tr>
<th>Fig. 5. (a) FDPs of MPEG-4 standard. (b) Proposed FDPs.</th>
</tr>
</thead>
</table>

| Fig. 6. (a) Original image with marked FDPs. (b) Corresponding caricature with marked FDPs. |
Figure 6 illustrates an example of original image-caricature pair with marked FDPs. The \((x, y)\) coordinate pairs of all manually marked feature points of the original face and the caricature are then forwarded to the next stage, i.e. mean face generation.

D. Mean Face Generation

The mean face is obtained by averaging the normalised original images of the training set (i.e. images 1 to 10 in figure 4), using the ‘Morpher [22]’ software. Original images 11 and 12 (see figure 4) are used as the validation set. After manually marking the feature points by using the approach described in subsection C, the \((x, y)\) co-ordinate pairs of these points can be easily captured by the ‘Morpher’ software package, which subsequently generates the mean of two input images (at a given time) using image morphing. Note that due to the limitation of ‘Morpher’ software to only use two input images at a given time, the averaging of the ten facial images has to be done, pair-wise. A simple morphing example that averages two faces using ‘Morpher’, is illustrated in figure 7.

In calculating the average of two facial images, the manually marked feature points play an important role. All the facial components, such as the mouth, nose, etc are first defined by the use of their FDPs. Subsequently, for each component, the feature points together are connected and the area enclosed is normalised, so that an area with equal texture is defined by each contour (see Figure 5(a)). In generating the mean face of two input images, the ‘Morpher’ averages the coordinates of the corresponding feature points and the colour values for each pixel. Figure 8(b) illustrates the final mean face generated from the ten original images.

E. Relationships among original image, corresponding caricature and mean face

Once the mean face has been generated, the procedure of marking and recording FDPs (as described in subsection C) will be performed on it. This data will then be used to estimate the deviation of each original image feature from the corresponding feature of the mean face. Figure 9 illustrates the relationships between the mean face, an original image and its corresponding caricature that is used in the proposed approach.

Even though the exact process of drawing a caricature from a facial image is hard to describe, previous research in psychology [3] has shown that it can be explained as follows: Every caricaturist has a mean face in his/her mind, which is a result of the human psycho-visual system working in a capacity similar to the ‘Morpher [22]’ explained in subsection D. This unconscious knowledge of the mean face gives the caricaturist the ability to point out distinctive features of a new facial image being viewed.

Let \(\Delta S\) be the difference between an original face, \(O\), and the mean face, \(M\). Therefore,

\[
\Delta S = O - M
\]  

(3)

Considering \(\Delta S\), the artist then exaggerates the original image to form a caricature. The difference between the caricature, \(C\), and its corresponding original image, \(O\), is the change made by the artist in drawing the caricature, which is defined as \(\Delta S'\).

\[
\Delta S' = C - O
\]  

(4)

In summary, when an artist sees the difference \(\Delta S\), then he/she makes the change \(\Delta S'\). Therefore the relationship between \(\Delta S\) and \(\Delta S'\) defines the artist’s drawing rules that govern his/her drawing style. It is known that different artists have different styles of drawing caricature as the rules embedded in their subconscious minds are different. Thus by capturing the relationship between \(\Delta S\) and \(\Delta S'\) from the drawings of an artist, the artist’s drawing style can be summarised into a set of rules. This provides the ability to apply the above captured rules to a totally new image, generating a caricature with this particular
has resulted in the design of several architectures that are always non-linear and the rules are difficult to describe in written language precisely. As a result, we adopt an artificial intelligence based approach, i.e. the use of a neural network, to accomplish this task (see subsection F).

F. Artificial Neural Network

An artificial neural network, commonly referred to as a "neural network", is a type of artificial intelligence technology that attempts to imitate the way a human brain works. It has been established that there are over 10 billion neurons in a human brain. These neurons are the basic units that provide humans with abilities to think, remember and to experience several sensations. Biological neurons are simulated by artificial neurons in artificial neural networks [23],[24]; they are the core elements that are connected in a particular logical manner to form an artificial neural network. Artificial neurons (simply called neurons hereafter) are usually connected in three layers. The first layer is an input layer, consisting of neurons that receive information (inputs) from the external environment. The second layer, which performs essential intermediate computations, is hidden from view (i.e. not directly visible from the external world) and is referred to as the hidden layer. The third layer is an output layer (target/output) that communicates the result of the weighted, summed output to the external environment or to the user. At the input layer, a linear input function computes the weighted sum of the inputs. Subsequently a non-linear transfer function transforms the weighted sum into final output values. Thus in general, all neural network architectures/topologies are based on the concept of input/output neurons, number of layers, a training function and transfer functions. Past research in neural network technology has resulted in the design of several architectures that are capable of solving specific problems.

The main reason for using neural networks in the proposed approach is their proven learning ability. A network is capable of learning from a training set by constructing an input-output mapping for the problem automatically. Therefore, an understanding of how the input is mapped to the output is not necessary, which is ideally suitable to be used for capturing the unexplainable relationship between ∆S and ∆S' (see subsection E) in caricature generation. Moreover, a neural network has the ability to capture non-linear relationships from a training set. The non-linear equations embedded in neurons are able to compute and provide results in a non-linear manner. This is suitable for capturing non-linear exaggerations of facial components in caricaturing.

The training set of a neural network consists of both input and output values. Therefore in the proposed approach, in order to capture the relationship between ∆S and ∆S', the input and output to the neural network should be ∆S and ∆S' respectively. Figure 10 illustrates a simple example that can be used to establish a relationship between ∆S and ∆S'. For the purpose of explanation, assume that each oval corresponds to a contour of an eye; each defined by eight feature points. The innermost oval is a mean eye, the one in the middle is an original eye and the outermost oval is a caricature eye. They are overlapped with each other by using the iris as the common reference point. In this example, the original eye is slightly bigger than the mean eye. Therefore we can expect the artist to exaggerate the caricature eye in a non-linear way, during which both the size and the shape of the eye can be changed.

Specifically defining ∆S to be the separation between two corresponding feature points of the original image (say O) and the mean shape image (say M); their separation in x and y directions can be written as:

\[ X_{AS} = X_O - X_M \]  \hspace{1cm} (5)
\[ Y_{AS} = Y_O - Y_M \]  \hspace{1cm} (6)

Similarly, defining ∆S' to be the separation between two corresponding feature points of the caricature (say C) and its corresponding original image, their separation in x and y directions can be written as:

\[ X_{AS'} = X_C - X_O \]  \hspace{1cm} (7)
\[ Y_{AS'} = Y_C - Y_O \]  \hspace{1cm} (8)

After calculating the ∆S(x,y) and ∆S'(x,y) of all feature points, a table consisting of the training set entries is prepared. Table II illustrates a section of this table. [Note: in our experiments the original image-caricature pairs, I to 10 in figure 4 have been used for training].

Table: 

<table>
<thead>
<tr>
<th>FDP No.(fig5)</th>
<th>X_{AS} of O-C Pair 1</th>
<th>X_{AS'} of O-C Pair 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11.1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>11.2</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>11.3</td>
<td>-4</td>
<td>-1</td>
</tr>
<tr>
<td>4.4</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

O-C = Original Image-Caricature

Once the training set has been prepared, the next step is to define a neural network. In this project, we propose the use of a feed-forward network with only one hidden layer as the architecture. The numbers of nodes in input, hidden and output
layers have been selected be the number of FDPs to be considered (see table I), i.e. 46. Table III tabulates the details of neural network architecture and training parameters used.

### TABLE III

<table>
<thead>
<tr>
<th>Architecture and Parameters</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Neural Networks</td>
<td>2 (x and y are trained separately)</td>
</tr>
<tr>
<td>Required per Artist</td>
<td></td>
</tr>
<tr>
<td>Neural Network Type</td>
<td>Feed-forward Back-propagation</td>
</tr>
<tr>
<td>Training Function</td>
<td>Levenberg-Marquardt</td>
</tr>
<tr>
<td>Performance Validation Function</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>Performance Goal</td>
<td>0</td>
</tr>
<tr>
<td>Minimum Gradient</td>
<td>1e-010</td>
</tr>
<tr>
<td>Maximum Number of Epochs</td>
<td>1000</td>
</tr>
<tr>
<td>Number of Layers</td>
<td>3</td>
</tr>
<tr>
<td>Hidden Layer Transfer Function</td>
<td>Tan-sigmoid</td>
</tr>
<tr>
<td>Output Layer Transfer Function</td>
<td>Pure-linear</td>
</tr>
<tr>
<td>Number of nodes in input layer</td>
<td>46</td>
</tr>
<tr>
<td>Number of nodes in hidden layer</td>
<td>46</td>
</tr>
<tr>
<td>Number of nodes in output layer</td>
<td>46</td>
</tr>
</tbody>
</table>

After the neural network is constructed, the training process can be commenced. The neural network was trained by using the Levenberg-Marquardt [29] algorithm without momentum. The mean squared error was used as the performance validation function and the performance goal was set to zero. The training stopped once the performance was minimized to the goal or the gradient of performance was less than the minimum gradient parameter, which means the error slope is close to zero and further training cannot reduce the error by much [29]. It would be unwise to design a network, train it and then put it into practice immediately. Its accuracy and capabilities should first be evaluated and scrutinized. This process is known as validation, which is important, as small errors could result in a misleading output from a network. The process of validation will be further discussed in section III. [Note: in our experiments the original image-caricature pairs 11 and 12 (see figure 4) are used for validation].

Once training and validation have been successfully completed as described above, the $\Delta S$ describing (at each FDP) the differences of a new original facial image from the mean face is fed as input to the trained neural network for testing. The neural network will then generate $\Delta S'$ that describes (at each FDP) the differences of predicted caricature image from the original image. As the FDPs of the original facial image are known, the above differences can be used to obtain the FDPs that define the corresponding caricature image, which are then forwarded to the final stage, mesh warping.

### G. Mesh Warping

Once the $(x,y)$ coordinate pairs defining the FDPs of both original image and the caricature to be generated are ready, they will enter a mesh warping module that converts the original image into its corresponding caricature. The idea of a mesh warping algorithm [25] is to deform one image (i.e. original) into another (i.e. caricature). We have specifically used the mesh warping algorithm of [25] for our experiments, though such algorithms are at present a common inclusion in most picture editing software packages [26],[27].

Within the proposed approach the $(x,y)$ coordinates of the FDPs of the original image form the source face mesh, and the XY coordinates of feature points generated from the neural network form the target face mesh. The warping module uses the source mesh and warps the original image towards the target mesh. Hence a caricature of the original face with the style of the particular artist (due to being warped towards the target mesh above) can be generated.

#### II.2 Component-Based Approach

![Component-based approach diagram](image)

In the entire face-based approach presented above, the use of a simple face model that uses 46 FDPs was proposed. However such a low number of FDPs is practically not sufficient to represent the detail and variety of a human facial figure. Nevertheless in provides means for maintaining the computational cost of the entire face-based approach at a relatively low level.

An obvious way to improve the accuracy of the entire face-based approach is the increase of the number of FDPs (resolution of the caricature), for e.g. to 145. Unfortunately this results in an exponential increase of the computational power and memory requirement of the neural network used, which is well beyond the capability of a typical desktop computer [Note: Specifications of the PC used in our experiments: Intel P4, 2.8GHz CPU, 1GB RAM]. An alternative solution to this problem is to consider the automatic caricaturing of individual facial components (such as nose, mouth, eyes etc.) with the total number of FDPs not exceeding 145 (say) and then joining the components subsequently to create the caricature of the entire face. We name this approach as the component-based
approach. It has the advantage of increasing the resolution of each individual component thereby increasing the overall accuracy of caricaturing process, yet maintaining the computational cost requirements.

In the first approach the advantage of training all facial components together is that the neural network is not only capable of learning the exaggerations of components but also on their relative positions. Therefore the facial components of the generated caricature are always positioned accurately relative to each other on the face. Unfortunately in the component-based approach, when considering components separately, the information on the relative positioning is lost. Therefore overlapping facial components exists in the resulting caricature if the generated components are combined together without a pre-processing stage. This commonly happens in cases such as when a heavily elongated nose overlaps with the mouth or exaggerated eyes overlap with unchanged eyebrows, due to their natural closeness. In order to remedy this problem we use the entire face-based approach as a means for creating a template for the exaggerated relative positioning of the facial components of the caricature. The individually caricatured facial components are then placed on the template forming the final caricature.

III. EXPERIMENTAL RESULTS AND ANALYSIS

Several experiments were carefully designed and carried out to prove the ability of the proposed approach to learn the drawing style of a caricaturist and automatically produce photorealistic caricature that is embedded with the unique drawing style of the artist. Three professional caricaturists were invited to draw caricatures of 12, male facial images from the AR face database (see section II.1.B for details and figure 4 for the results). MATLAB and its ANN toolbox were used as the programming language/environment [29]. Due to the relatively small number of validation images, a cross validation strategy in which, ‘10 out of the 12 images were used for training the neural network and the remaining two were used for validation’, was adopted. The results presented in this paper is a single instance of this cross validation, i.e., the case in which the original image-caricature pairs numbered 1-10 in figure 4, have been used for training and image-caricature pairs 11 and 12 as the validation set. In order to further reduce the need for excessive computer resource during the neural network training process, X and Y coordinates were trained separately in two identical neural networks, which were constructed based on the parameters provided in section table III.

Subsequent to the training of the neural network (see section II.1.F) using original image-caricature pairs 1-10, the ΔS'(x,y) of each FDP of image-caricature pairs 11 and 12 (i.e. validation images), obtained from the trained neural network's output are converted to (x,y) coordinates and forwarded to the mesh warping module (see section II.1.G) for the creation of the photorealistic caricatures, which are then compared with caricaturist's drawings for validation.

The experimental results for the three artists named artist-1, artist-2 and artist-3 are separately presented in figure 12, 13 and 14.
14 respectively. These figures consist of the original images which are illustrated in the first column, the caricatures of the original faces drawn by the artists in the second column, the automatically generated caricatures by the proposed entire face-based approach in the third column and the caricatures generated by the proposed component-based approach in the fourth column respectively. Note that the first and second rows separate the results for the two validation images.

The analysis of the results were done using two different approaches. The first is a thorough visual analysis of the caricatures by the authors who have significant understanding of the caricaturing process and the strengths and weaknesses of the existing and proposed approaches to automatic caricature generation. The second was based of presenting the results to 46 subjects, who are novice to caricaturing, but would be able to comment on the quality of the output results based on a general visual analysis. The evaluations based on the above approaches can be presented as follows:

**Author Analysis:** A careful comparison of columns 1 (original) and 2 (caricature drawn by the artist) of figure 12 (represents the artist-I’s caricatures) and column-1 of figure 4 reveals that in general there appears to be several traits in his drawing style. Two of these traits can be listed as follows:

a) In the horizontal direction the forehead area is exaggerated less as compared to the area near the cheek bone. Thus forehead area appears narrower as compared to cheek bone area.

b) The unique facial features have been significantly exaggerated as compared to what has been done by the other two artists.

[Note: the identification of the above drawing styles/traits requires the careful comparison of the above artist’s caricatures to the caricatures of the two other artists illustrated in figure 13 and 14.]

A visual analysis of the caricatures generated by the proposed approach for the two validation images illustrated in figure 12 reveal that the above drawing traits of the artist-I have been maintained by the proposed approach. In the computer generated caricatures of figure 12 (column 3), the forehead area is narrower as compared to the cheek bone area. Specifically the computer generated caricatures of artist-I corresponds more to the hand drawn caricature in terms of the above traits, than in the case of the other two artists. This is due to the fact that artist-I’s drawing style appears to be partly defined by the above trait (see figure 4).

Similar traits can be identified in the caricatures drawn by the artists 2 (fig. 13 and column 3 of fig. 4) and 3 (fig. 14 and column 4 of fig. 4). For example, the artists-2 appears to be elongating faces in the vertical direction in caricaturing whilst the artist-3 pays significant attention to the details of the eyebrows. The caricatures of these artists generated by the proposed approach illustrated in columns 3 in figures 13 and 14 prove that these drawing traits have been maintained. In general artist-I’s drawing style appears to have the most significant exaggerations and artist-3’s, the least. A careful comparison of figures 12, 13 and 14 reveals that computer generated images of artist-I’s caricatures have the highest similarity to the hand-drawn caricatures, while artist-3 has the least.

When comparing the component-based approach (column-4) to the entire face-based approach (column-3), component-based approach appears to be more successful in capturing the drawing style. This is expected due to the reasons discussed in section 11.2.

**Subjective Analysis:** Two subjective tests were carried out to further investigate our claims. 46 volunteers who have little or no knowledge about the art/science of caricaturing were used in the study.

In the first subjective test, a questionnaire was designed to investigate the ability of the proposed approaches in general to capture the drawing style of an artist and hence automatically reproduce caricatures with the same style. All subjects were shown pictures of two original faces, hand drawn caricatures of the face by the three artists (with ownership assigned) and the computer generated caricatures when the entire face-based approach was used, without the ownership being assigned. The task of the subjects was to select the owner of each unclassified computer generated caricature. The tests were performed for both validation images. The results are represented in figure 15 [Note that case 1 and 2 represent the corresponding computer generated caricatures 11 and 12 of that particular artist respectively]. The results show that more than 75% of the subjects were able to correctly determine the ownership of each computer generated caricature when using the entire face-based approach. Due to the unique vertical elongation drawing style adopted by the artist-2, his computer generated caricatures have been identified more accurately, on the average, with artist-I’s caricatures coming a close second (due to acute feature exaggerations used) and artist-3’s, last (due to minimal exaggerations used in his drawing style).

The subjective test results obtained in the first test were further analysed using statistical decision theory [30]. The hypothesis that 'subjects have been able to correctly match the artists caricature with that of the computer generated caricature using the proposed algorithm', was found to hold at a confidence level reaching 100%, for all images drawn by all three artists.

![Percentage of participants matching the caricatures correctly](fig15)

**Fig. 15. Results of subjective test 1.**
In the second subjective test, a different questionnaire was designed to compare the effectiveness of the entire face-based and component-based approaches. For each caricaturist and for each validation image, 46 subjects were provided with the original image, the caricature drawn by the artist and the photorealistic caricatures generated from two different approaches. The computer generated caricatures were presented in a mixed manner, without stating the approach that was followed to obtain it, i.e. whether the entire or component-based approach was used. The subjects were asked to decide which of the two computer generated images were more similar to the caricature drawn by the artist. This questionnaire gives a more direct comparison between two approaches, as compared to repeating the first subjective test on the component-based approach. The results are presented as a bar chart in figure 16. The bar chart represents the percentage of participants who found that the component-based approach was better than the entire face-based approach. For all three artists and for both validation cases, it is clearly illustrated that the component-based approach provides much superior quality caricatures. Therefore it can be concluded that the component-based approach has a better ability to capture the drawing style of an individual caricaturist as compared to the entire face-based approach. This justifies the claims made in section II.2.

The subjective test results obtained in the second test were also further analysed using statistical decision theory [30]. The hypothesis that 'the component-based approach was better than the entire face-based approach', was found to hold at a confidence level above 99.68%, for both validation cases of all three artists.

The two validation images are also exaggerated using the Picasso algorithm [8] as benchmarks for comparisons. Figure 17 illustrates the caricatures of the validation images produced when using the Picasso method with exaggerations to each feature done by multiplying the original facial image's corresponding feature difference from the mean by a factor of 1.5. Whilst this approach produces a caricature it does not produce caricatures that imitate the drawing style of a given artist, due to the fact that no drawing style capture algorithm has been used in the design.

![Fig. 17. Caricatures produced by the Picasso benchmark with ΔS* = 1.5 x ΔS.](image)

### IV. PRACTICAL CHALLENGES & APPLICATIONS

In section III we provided experimental results and a detailed analysis to prove that the proposed example-based neural network approach to automatic caricature generation is able to capture, learn and re-produce the drawing style of a caricaturist. Though this proof was based on testing a limited set of original image-caricature pairs drawn by a limited amount of caricaturists, the significantly good experimental results obtained and the detailed analysis carried out, justifies the claims made. However training on a larger set of original image-caricature pairs (belonging to different people but caricatured by the same artist) will help improve prediction further. Unfortunately, requesting a caricaturist to draw a large number of caricatures of a large database of facial images is challenging and somewhat unreasonable. Besides, the size of the training set is another considerable issue when computational cost is considered for processing and storage. Added to the above challenges, a large number of FDPs have to be defined if one is to expect a higher quality of caricature generation. However this increases the need of computational resources for processing. After careful consideration of the above mentioned practical limitations of the experimentation process, we have adopted a cross validation scheme for analysing the performance of the proposed system and have proposed the use of a component-based approach in which a larger amount of FDPs can be defined on a per-component basis.

The proposed scheme has potential to be applied in many areas. Though generating a caricature from a facial image is considered a popular art, the ability of a computer based system to learn the drawing style of an artist plays a significant role in its possible application domains. One such application domain is the use of the proposed approach to reproduce caricatures that have embedded styles of late, historically famous caricaturists. A busy caricaturist can make use of the proposed system to provide assistance in generating copyrighted computer generated caricatures, embedding his/her own drawing style. Indirectly the drawing style capture method proposed here can be used in more advanced computer vision applications such as, copyright theft prevention, assistance during legal proceedings, person identification via police drawings, etc.
example-based neural network approaches that are able to caricature generation. At present we are considering further caricatures drawn by an artist. The trained neural network is capturing the drawing style of artists from different cultures, capture and identify the drawing style of an artist by training a different genders project.

automatic caricature generation systems in their inability to and analytical experimental results have been provided to prove the possible effective use of the algorithms in automatic caricature generation. At present we are considering further improvement of the proposed approaches via providing means for fully automatic feature extraction, caricaturing faces of

ACKNOWLEDGMENT

We would like to thank Mr. Abuhelga, Mr. Yang and Mr. Pang for their contributions in drawing caricatures for the project.

REFERENCES
