由有複雜背景的連續影像中自動即時產生會說話卡通臉之研究及其應用

研究生：陳彥龍　　指導教授：蔡文祥 博士

國立交通大學資訊科學與工程研究所

摘要

本論文提出了一套自動即時產生會說話卡通臉的系統。這個系統包含了五個階段：環境學習、臉部特徵追蹤、影像特徵點轉換、錄音與動畫製作。在環境學習階段，系統會自動學習一些特徵值來輔助臉部特徵追蹤。在臉部特徵追蹤階段，系統會由連續影像追蹤眼睛和嘴形的變化，我們提出了一些臉部特徵追蹤的方法來達到此結果，除此之外更提出了一個能夠偵測轉頭的方法。在我們提出的錯誤校正方法輔助之下，臉部特徵追蹤的方法也可用於晃動的人臉上。在影像特徵點轉換階段，我們提出一個方法將追蹤到的特徵點參數轉換成相對應的人臉模型的控制點。在錄音階段，聲音將會持續被錄音及即時播放出來。在動畫製作階段，我們根據人臉模型控制點的位置產生卡通臉，並提出一個方法產生側面的二維卡通臉動畫。以這套系統為基礎，我們發展出另一套網路應用的自動即時產生會說話卡通臉的系統，並實作出二種有趣的應用。實驗結果證實本論文所提出方法之可行性。

Automatic Real-time Generation of Talking Cartoon Faces from Image Sequences in Complicated Backgrounds And Applications

Student: Yeng-Long Chen Advisor: Dr. Wen-Hsiang Tsai

Institute of Computer Science and Engineering
National Chiao Tung University

ABSTRACT

A system for automatic real-time generation of talking cartoon faces is proposed, which includes five processes: environment learning, facial feature tracking, image feature point transformation, speech recording, and animation generation. In the environment learning process, some threshold values are learned to help tracking facial features. In the facial feature tracking process, the eyes and mouths are tracked from sequential facial images. An eye-pair tracking method and a mouth tracking method are used to reach the goal. Besides, a head turning detection method is proposed. Some error correction techniques are proposed to cause the facial tracking methods be applicable to shaking faces. In the image feature point transformation process, a method used to transform image feature points into 2D face model control points is proposed. In the speech recording process, speeches are recorded and played in real time. In the animation generation process, talking cartoon faces are rendered from the face model control points. A method for creation of lateral 2D cartoon faces is also proposed. Based on the proposed system, an automatic real-time talking cartoon face generation system for use on networks is designed. Two kinds of interesting applications on networks are also implemented. Experimental results show the feasibility of the proposed methods.

ACKNOWLEDGEMENTS

The author is in hearty appreciation of the continuous guidance, discussions, support, and encouragement received from his advisor, Dr. Wen-Hsiang Tsai, not only in the development of this thesis, but also in every aspect of his personal growth.

Thanks are due to Mr. Tsung-Yuan Liu, Mr. Chih-Jen Wu, Mr. Kuo-Feng Chien, Miss Pei-Pei Chen, Mr. Kai-Li Chiang, Miss Yu-Tzu Wang, Miss Chia-Yu Hsu, and Miss Yi-Lin Wang for their valuable discussions, suggestions, and encouragement. Appreciation is also given to the colleagues of the Computer Vision Laboratory in the Institute of Computer Science and Engineering at National Chiao Tung University for their suggestions and help during his thesis study.

Finally, the author also extends his profound thanks to his family for their lasting love, care, and encouragement. He dedicates this dissertation to his parents.

CONTENTS

ABSTRACT (in Chinese) i
ABSTRACT (in English) ii
ACKNOWLEDGEMENTS iii
CONTENTS iv
LIST OF FIGURES vii

Chapter 1 Introduction 1
1.1 Motivation 1
1.2 Survey of Related Studies 2
1.3 Overview of Proposed Method 4
1.3.1 Definitions of Terms 4
1.3.2 Assumptions 5
1.3.3 Brief Descriptions of Proposed Method 6
1.4 Contributions 8
1.5 Thesis Organization 9

Chapter 2 Cartoon Face Modeling from Image Feature Point Transformation 10
2.1 Introduction 10
2.2 Review of Adopted Cartoon Face Model 11
2.3 Transformation of Image Feature Points into Face Model Control Points 14
2.3.1 Transformation Technique 14
2.3.2 Transformation Involving Eye Feature Points 14
2.3.3 Transformation Involving Mouth Feature Points 16
2.4 Creation of Cartoon Face 18
2.4.1 Creation of Frontal Cartoon Face 19
2.4.2 Creation of Lateral Cartoon Face 22
2.5 Experimental Results 24

Chapter 3 Facial Feature Tracking from Image Sequences in Complicated Backgrounds 27

3.1 Introduction 27
3.2 Review of Employed Tracking Method 28
3.3 Segmentation of Facial Feature Regions 30
3.3.1 Segmentation of Eye Regions 30
3.3.2 Segmentation of Mouth Regions 32
3.4 Proposed Method of Eye-pair Tracking 34
3.4.1 Extraction of Eye-pair Regions 35
3.4.2 Correction of Region Extraction Errors 36
3.5 Proposed Method of Mouth Tracking 38
3.5.1 Extraction of Mouth Regions 39
3.5.2 Correction of Region Extraction Errors 42
3.6 Detection of Head Turning 43
3.6.1 Basic Idea 44
3.6.2 Detection Process 44
3.7 Experimental Results 46

Chapter 4 Automatic Generation of Talking- Face Cartoon Videos 49
4.1 Introduction 49
4.2 Overview of AVI Video File Format 50
4.2.1 Overview of RIFF Format 50
4.2.2 Overview of AVI Format 51
4.3 Generation of Cartoon Videos from AVI Files 53
4.3.1 Basic Idea and Configuration 53
4.3.2 Generation Process 55
4.4 Generation of Cartoon Videos from Face Feature Data in Processed Image Frames 55
4.4.1 Basic Idea and Configuration 56
4.4.2 Generation Process 57
4.5 Experimental Results 57

Chapter 5 Real-time Cartoon Face Animation 61
5.1 Introduction 61
5.2 Learning of Environment 63
5.2.1 Learning of Eye-pair Region Threshold Value 63
5.2.2 Learning of Division Line 64
5.2.3 Learning of Mouth Region Threshold Value 65
5.3 Real-time Cartoon Face Generation Process 66
5.4 Speech Recording and Play 68

5.4.1 Overview of WAVE Audio File Format 68
5.4.2 Process of Recording and Play 70
5.5 Synchronization of Cartoon Videos and Speeches 71
5.6 Experimental Results 73

Chapter 6 Applications of Real-time Cartoon Face Animation to Multi-role Avatar Broadcasting and Web TV through Networks 77
6.1 Introduction 77
6.2 Server and Client System for Animation Broadcasting 78
6.2.1 System Organization 78
6.2.2 Video and Audio Transmission 79
6.3 Synchronization of Cartoon Videos and Speeches 81
6.4 An Application to Multi-role Avatar Broadcasting through Networks 82
6.4.1 Idea 82
6.4.2 Process of Creation 83
6.5 An Application to Web TV by ActiveX Technique 84
6.5.1 Idea 84
6.5.2 Process of Creation 85
6.6 Experimental Results 86

Chapter 7 Conclusions and Suggestions for Future Works 93
7.1 Conclusions 93
7.2 Suggestions for Future Works 94

References 96

LIST OF FIGURES

Figure 1.1 A flowchart of talking-face cartoon video generation system. 6
Figure 1.2 A flowchart of real-time talking cartoon face generation system 7
Figure 1.3 A flowchart of real-time talking cartoon face generation system through networks 8
Figure 2.1 An overall configuration of proposed talking cartoon face generation system 11
Figure 2.2 Flowchart of hierarchical bi-level thresholding method in Chen and Tsai [1]. 12
Figure 2.3 A face model. (a) Proposed 72 feature points. (b) Proposed facial animation parameter units in Chen and Tsai [1]. 13
Figure 2.4 An illustration of corner-cutting algorithm in Chen and Tsai [1]. 13
Figure 2.5 Cubic Bezier curve in Chen and Tsai [1]. 13
Figure 2.6 An illustration of image feature point transformation for eyes..... 15
Figure 2.7 The open range of the upper lip. 17
Figure 2.8 An illustration of image feature point transformation for mouth. 18
Figure 2.9 Points to help drawing. 18
Figure 2.10 An illustration of arc(P1, …, Pn). 19
Figure 2.11 An illustration of the steps in the creation of frontal cartoon face. (a) The creation of the contour of a face. (b) The creation of the ear. (c) The creation of the nose. (d) The creation of the eyebrow. (e) The creation of the eye. (f) The creation of the mouth. 22
Figure 2.12 An experimental result of the creation of frontal cartoon face. (a) A male face model. (b) A female face model. 22
Figure 2.13 An illustration of creation of lateral cartoon face. 23
Figure 2.14 An illustration of creation of lateral cartoon faces. (a) A lateral cartoon face with t = 8. (b) A lateral cartoon face with t = $\tilde{}$ 8. 24
Figure 2.15 An example of experimental results for fontal cartoon faces. (a)(d)(g)(j) The input sequential facial image. (b)(e)(h)(k) The first result of cartoon face (c)(f)(i)(l) The second result of cartoon face 25
Figure 2.16 An example of experimental results for lateral cartoon faces (a)(d)(g)(j) The input sequential facial image. (b)(e)(h)(k) The first result of cartoon face (c)(f)(i)(l) The second result of cartoon face 26
Figure 3.1 Flowchart of proposed method for facial feature tracking..... 28
Figure 3.2 A flowchart of talking cartoon face generation from image sequences in Chen and Tsai [1] 29
Figure 3.3 An experimental result of segmentation of eye-pair regions in intensity

channel. (a) The neutral facial image in intensity channel (b) The binary image.....32
Figure 3.4 An illustration of the octagon fitting in the rectangle. (a) The composing points in the octagon. (b) The octagon in a binary image.....33
Figure 3.5 An experimental result of segmentation of mouth regions. (a) The neutral facial image. (b) The resulting binary image. 34
Figure 3.6 An experimental result of extraction of eye-pair regions. (a) The binary image Beye. (b) The square in Beye with n = 9. (c) The final result of extraction of eye regions. 36
Figure 3.7 A flowchart of eye-pair tracking method. 38
Figure 3.8 An illustration of rectangle Rmouth. 40
Figure 3.9 An experimental result of preprocessing for extraction of mouth regions. (a) An illustration of mouth region. (b) The preprocessing result of mouth region extraction. 40
Figure 3.10 An experimental result of extraction of mouth regions. (a) The original image. (b) Final result of mouth region tracking 41
Figure 3.11 A flowchart of proposed mouth tracking method 43
Figure 3.12 Some definitions for detection of head turning.....44
Figure 3.13 An experimental result of detection of head turnings. (a) The Sobel edge image E (b) The x-position of the left and right cheeks and the center of the mouth. 46
Figure 3.14 An experimental result of eye-pair region tracking in sequential facial images. 47
Figure 3.15 An experimental result of mouth region tracking in sequential facial images. 47
Figure 3.16 An experimental result of detection of head turnings in sequential facial images. 48
Figure 4.1 An illustration of an RIFF format. 51
Figure 4.2 An illustration of an AVI format. 52
Figure 4.3 A Configuration of the talking-face cartoon video generation system from AVI files 54
Figure 4.4 A configuration of talking-face cartoon video generation system from face feature data 56
Figure 4.5 An input video sequence used for the experiment. 58
Figure 4.6 A resulting sequence of the talking-face cartoon video for “夢想”..59
Figure 4.7 Another resulting sequence of the talking-face cartoon video for “夢想” 60
Figure 5.1 A configuration of the system for real-time generation of talking cartoon

faces. 62
Figure 5.2 An illustration of B1 and B2. 64
Figure 5.3 A flowchart of the process for real-time generation of cartoon faces. 67
Figure 5.4 An illustration of WAVE PCM files 69
Figure 5.5 A flowchart of the system for real-time talking speeches. 71
Figure 5.6 A flowchart of the process for synchronization of videos and audios. 73
Figure 5.7 The program interface of real-time talking cartoon face generation. 74
Figure 5.8 The process of learning of environments before starting real-time facial feature tracking. 74
Figure 5.9 The interface for the process of choosing a face model 75
Figure 5.10 An experimental result of real-time frontal talking cartoon faces. 75
Figure 5.11 An experimental result of real-time lateral talking cartoon faces. 76
Figure 6.1 A configuration of the real-time cartoon face animation system. 79
Figure 6.2 The structure of the video unit. (a) A header unit. (b) A frame unit 80
Figure 6.3 The structure of the audio unit. 80
Figure 6.4 A flowchart of the process for synchronization of videos and audios. 82
Figure 6.5 An illustration of multi-role avatar broadcasting. 83
Figure 6.6 The block diagram of creating a multi-role avatar broadcasting. 84
Figure 6.7 An illustration of the web TV. 85
Figure 6.8 The block diagram of creating a web TV. 85
Figure 6.9 The program interface of multi-role avatar broadcast system in the server site. 87
Figure 6.10 The program interface of multi-role avatar broadcast system in the client site. 87
Figure 6.11 The process of learning of environments in the server site. 88
Figure 6.12 The process of waiting for the connections from the client site. 88
Figure 6.13 The choice of the face model in the client site 89
Figure 6.14 The process of connecting to the server in the client site. 89
Figure 6.15 The real-time process of facial feature tracking in the server site. 90
Figure 6.16 The resulting real-time talking cartoon face in the client site. 90
Figure 6.17 The process of entering the URL of the server and downloading the ActiveX control from the server. 91
Figure 6.18 The program interface of web TV in the Web browser. 91
Figure 6.19 The resulting real-time talking cartoon face in Web browser. 92