AUTOMATIC RECOGNITION OF FACIAL EXPRESSIONS USING HIDDEN MARKOV MODELS AND ESTIMATION OF EXPRSSION INTENSITY

by

Jenn-Jier James Lien

B.S. in Biomedical Engineering, Chun Yuan University, Taiwan, 1989

M.S. in Electrical Engineering, Washington University, St. Louis, Missouri, 1993

Submitted to the Graduate Faculty

of the School of Engineering

in partial fulfillment of

the requirements for the degree of

Doctor

of

Philosophy

1998

COMMITTEE SIGNATURE PAGE

This dissertation was presented

by

Jenn-Jier James Lien

It was defended on

April 14, 1998

and approved by

(Signature) ______________________________
Committee Co-Chair
Dr. Ching-Chung Li, Professor of Electrical Engineering

(Signature) ______________________________
Committee Co-Chair
Dr. Takeo Kanade, Professor and Director of the Robotics Institute,
School of Computer Science, Carnegie Mellon University

(Signature) ______________________________
Committee Co-Chair
Dr. Jeffrey F. Cohn, Professor of Psychology

(Signature) ______________________________
Committee Member
Dr. Henry Y.H. Chuang, Professor of Computer Science

(Signature) ______________________________
Committee Member
Dr. Richard W. Hall, Professor of Electrical Engineering

(Signature) ______________________________
Committee Member
Dr. Morton Kanefsky, Professor of Electrical Engineering

(Signature) ______________________________
Committee Member
Dr. Marwan A. Simaan, Professor and Chairman of Electrical Engineering

ACKNOWLEDGEMENTS

I am greatly indebted to Professor Takeo Kanade and Professor Ching-Chung Li, my co-advisors and my mentors, for not only providing invaluable guidance, advice, criticism and encouragement but also giving me the latitude I have needed to develop as a researcher. I thank Professor Jeffrey Cohn, co-advisor, for his support and teaching with the Facial Action Coding System (FACS). I would also like to thank the other members of my thesis committee: Professors Richard Hall, Morton Kanefsky, Marwan Simaan and Henry Chuang for their valuable suggestions and feedback.

My years at the Vision and Autonomous Systems Center of the Robotics Institute, Carnegie Mellon University have been priceless. I consider myself lucky to be a part of this research center and spend many nights with friends discussing technical issues. I would like to thank Jie Yang and Michael Nechyba for sharing their experiences with Hidden Markov Models; Conrad Poelman, Richard Madison and Yalin Xiong for exchanging knowledge on optical flows; Peter Rander, Henry Rowley, Shumeet Baluja, Teck Khim, Wei Hua, Mei Han, Mei Chen, Dongmei Zhang, Michael Smith, Daniel Morris and Farhana Kagalwala for much needed help; and Adena Zlochower for her help with FACS on the facial expression analysis project. My thanks also go to Chung-Hui Anne Lin for her support and encouragement. I would especially like to thank David LaRose and Yu-Te Wu, who have been tremendous fun to work and play with, and who have provided countless hours of invaluable discussions on the topics presented here.

Special thanks to Matthew Turk, Steve Shafer, P. Anandan, Richard Szeliski and Harry Shum at Microsoft Vision group for their valuable comments and suggestions.

Finally, I would like to thank my parents, Chin-Chuan Lien and Wen-Hua Shih, as well as my other family members: Jenn-Ren Lien, Jenn-Yueh Lien, Shu-Hua Chien, Hui-Yin Christia Tien for their constant love, support and encouragement. Without them, none of this would have been possible. I cannot begin to thank them enough, and they will always have my respect and love.

ABSTRACT

Signature ________________________
Professor Ching-Chung Li

Signature ________________________
Professor Takeo Kanade

Signature ________________________
Professor Jeffrey F. Cohn

AUTOMATIC RECOGNITION OF FACIAL EXPRESSIONS USING HIDDEN MARKOV MODELS AND ESTIMATION OF EXPRSSION INTENSITY

Jenn-Jier James Lien, Ph.D.

Facial expressions provide sensitive cues about emotional responses and play a major role in the study of psychological phenomena and the development of nonverbal communication. Facial expressions regulate social behavior, signal communicative intent, and are related to speech production. Most facial expression recognition systems focus on

only six basic expressions. In everyday life, however, these six basic expressions occur relatively infrequently, and emotion or intent is more often communicated by subtle changes in one or two discrete features, such as tightening of the lips which may communicate anger. Humans are capable of producing thousands of expressions that vary in complexity, intensity, and meaning. The objective of this dissertation is to develop a computer vision system, including both facial feature extraction and recognition, that automatically discriminates among subtly different facial expressions based on Facial Action Coding System (FACS) action units (AUs) using Hidden Markov Models (HMMs).

Three methods are developed to extract facial expression information for automatic recognition. The first method is facial feature point tracking using the coarse-to-fine pyramid method, which can be sensitive to subtle feature motion and is capable to handle large displacements with subpixel accuracy. The second is dense flow tracking together with principal component analysis, where the entire facial motion information per frame is compressed to a low-dimensional weight vector for discrimination. And the third is high gradient component (*i.e.*, furrow) analysis in the spatio-temporal domain, which exploits the transient variance associated with the facial expression.

Upon extraction of the facial information, non-rigid facial expressions are separated from the rigid head motion components, and the face images are automatically aligned and normalized using an affine transformation. The resulting motion vector sequence is vector quantized to provide input to an HMM-based classifier, which addresses the time warping problem. A method is developed for determining the HMM topology optimal for our recognition system. The system also provides expression intensity estimation, which has significant effect on the actual meaning of the expression.

We have studied more than 400 image sequences obtained from 90 subjects. The experimental results of our trained system showed an overall recognition accuracy of 87%, and also 87% in distinguishing among sets of three and six subtly different facial expressions for upper and lower facial regions, respectively.

DESCRIPTORS

Action Unit (AU)
Computer Vision and Pattern Recognition
Dense Flow
Eigenflow
Expression Intensity Estimation
Facial Action Coding System (FACS)
Facial Expression Recognition
Feature Point Tracking
Furrow Extraction
Hidden Markov Model (HMM)
Human-Computer Interaction (HCI)
Motion Line/Edge Extraction
Optical Flow
Principal Component Analysis (PCA)
Rigid and Non-Rigid Motion
Wavelet-Based Motion Estimation

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS…..…………………………………………….………………… iii

ABSTRACT………………………………………..…………..………...…… iv

LIST OF FIGURES…...………………………………………..……………..… x

LIST OF TABLES….………………………………………………….……….xvii

1.0 INTRODUCTION……..………………………………………………...….. 1

1.1 Related Works…...…………………………………..…………….….. 2

1.2 Problem Statement..………………………………………….…………… 7

1.3 Objective of the Research.……………………………………...… 10

1.4 Organization of the Dissertation….……………………………………. 11

2.0 FACIAL EXPRESSION RECOGNITION SYSTEM OVERVIEW…………. 13

2.1 Three Methods of Feature Motion Extraction...……………………….. 13

2.2 Recognition Using Hidden Markov Models..…………………….… 16

2.3 Facial Action Coding System and "Expression Units".……………….. 18

2.4 Rigid and Non-Rigid Motion Separation and Geometric Normalization………………………………………..………… 22

3.0 FACIAL FEATURE POINT TRACKING…………………………….…. 25

3.1 Dot Tracking and Reliability of Feature Point Selection..……….……. 25

3.2 Motion Estimation and Flow Window…………………………………. 30

3.3 Motion Confidence Estimation.……………………………...… 35

3.4 Tracking Subpixel and Large Motion………………………………….. 37

3.5 Analysis of Feature Point Tracking Problems…………….………… 42

3.6 Data Quantization and Conversion for the Recognition System………. 43

3.7 Expression Intensity Estimation..……………………………….... 49

4.0 DENSE FLOW TRACKING AND EIGENFLOW COMPUTATION……….. 53

4.1 Wavelet-Based Motion Estimation…………………………………….. 53

4.2 Dense Flow Tracking…….…………………………………………… 60
4.3 Eigenflow Computation..………………………………………….... 62
4.4 Data Quantization and Conversion for the Recognition System……..… 70
4.5 Correlation and Distance in Eigenspace……………………………….. 77
4.6 Expression Intensity Estimation……...…………………………….... 78
5.0 HIGH GRADIENT COMPONENT ANALYSIS……..……………………… 80
5.1 High Gradient Component Detection in the Spatial Domain………….. 80
5.2 High Gradient Component Detection in the Spatio-Temporal Domain……………………………………………………………. 81
5.3 Morphology and Connected Component Labeling……………………..… 86
5.4 Analysis of High Gradient Component Detection Problems………...… 96
5.5 Data Quantization and Conversion for the Recognition System………. 98
5.6 Expression Intensity Estimation.………………………………………. 102
6.0 FACIAL EXPRESSION RECOGNITION USING HIDDEN MARKOV MODELS…………………………..…………………………….....…104
6.1 Preprocessing of Hidden Markov Models: Vector Quantization……..… 105
6.2 Beginning from Markov Models………………………………..…… 109
6.3 Extension of Markov Models: Hidden Markov Models.………………. 111
6.4 Three Basic Problems of Hidden Markov Models…………………..… 114
6.4.1 Probability Evaluation Using the Forward-Backward Procedure………………………………………………………114
6.4.2 Optimal State Sequence Using the Dynamic Programming Approach…………………………………………..…………121
6.4.3 Parameter Estimation Using the Baum-Welch Method………... 124
6.5 Computation Considerations..…………………………………..……128
6.5.1 Choice of Hidden Markov Model..…………………………….. 128
6.5.2 Initialization of Hidden Markov Model Parameter Estimation…………………………………………………… 131

6.5.3 Computation of Scaling…..…………………………...……….. 131
6.5.4 Computation of Smoothing for Insufficient Training Data……. 134
6.5.5 Computation of Normalization…..…………………………….. 135
6.5.6 Computation of Convergence……………………………..…… 136
6.5.7 Computation of Confidence…..…………………………….... 137
7.0 DETERMINATION OF HIDDEN MARKOV MODEL TOPOLOGY……..… 138
7.1 The Method…………………………………………………..……138
7.1.1 Step 1: The 1st-Order Markov Model…………..……………... 139
7.1.2 Step 2: The 1st-Order Hidden Markov Model……………..…...140
7.1.3 Step 3: The Multi-Order Hidden Markov Model……………….159
7.2 Physical Meaning of Hidden Markov Model Topology..………..……. 162
8.0 EXPERIMENTAL RESULTS.…………………………………………..… 164
8.1 Data Acquisition, Experimental Setup, and Digitizing……………...… 164
8.2 Segmentation and Coding by Human Observers (Ground Truth)…….. 166
8.3 Automated Expression Recognition.…………………………………... 171
8.3.1 Training Process..…………...…………………………………..… 171
8.3.2 Recognition Results………………….………………………… 178
9.0 CONCLUSIONS……………………………………..…………..…..... 187
9.1 Contributions…………………….……...…………………………... 187
9.2 Suggestions for Future Work………………...………………………....189
APPENDIX………………………………………………………………….…191
BIBLIOGRAPHY……………………………………………………………....196

LIST OF FIGURES

Figure No. Page

1 Comparison of different smile expressions with different expression intensities. The presence or absence of one or more facial actions can change their interpretations………………………………………………………………….. 8

2 The user interface created by programming in C, Motif, X Toolkit and Xlib…. 14

3 Block diagram of a facial expression recognition system……………………… 15

4 “Expression units” of subtly different facial expressions in our study (taken from [34])……………………………………………………………………... 19

5 Normalization of each face image to a standard 2-dimensional face model…… 23

6 Dot tracking: each dot is marked by a cross (+) at its center, lines trailing from the dots represent changes in the location of dots due to facial expression…... 28

7 Locations of selected facial feature points (marked by a cross ‘+’) which reflect the muscle motion of facial features……………………………………………. 29

8 Feature point tracking based on tracking the movement of an n x n feature window between two consecutive frames (here, n is 13 pixels)……………….. 30

9 Window (weight) function $w(x)$ can be used to control the accuracy of motion estimation based on the gradient varying from point to point [66]……………... 33

10 Comparison of Lucas-Kanade, spline-based and wavelet-based window functions…………………………………………………………………………... 34

11 Trackability of various feature regions in a binary image: (a) contains no image texture so it would make a poor feature; (b) and (c) have high gradients locally either in one direction, or the horizontal- (I'_c-) and vertical- (I'_r-) gradients are highly correlated with each other as in (c), they also are not trackable; only feature (d) can be used as a trackable feature [77]…………….……………..…. 36

12 Feature point tracking excluding the pyramid method: it is sensitive to subtle motion such as eye blinking, but it loses tracking for large motion such as

mouth opening and suddenly raising eye brows. Lines trailing along feature points (marked by a cross '+') denote their movements across image frames in the sequence………………………………………………………………. 39

13 A 5-level pyramid for feature point tracking…………………………………... 40

14 Feature point tracking including the pyramid method: it is sensitive to subtle motion such as eye blinking and also tracks accurately for large motion such as mouth opening and suddenly raising eye brows……………………………... 44

15 The feature point tracking error due to violation of the feature region's trackability condition: tracking along the edge direction at both brows and mouth regions with deformed shapes…………………………………………... 45

16 Reducing the tracking error by locating feature points away from edges of facial features: the tracking for brow region is improved, but is still erroneous in mouth region with large mouth opening………………………………….. 45

17 Reducing the tracking error by using a large window size which requires more processing time………………………………………………………………… 46

18 Reducing the tracking error by increasing the value of the weighting factor *K* (*K*=0.5) of the inter-level confidence base in order to include more global information from the previous low-resolution level processing in the pyramid method……………………………………………………………………. 46

19 Displacement vector of the facial feature point tracking to be encoded for input to a Hidden Markov Model…………………………………………………..… 47

20 An example illustrating the expression intensity estimation by following the non-linear mapping (k = 7) of the constrained feature point displacement for the case of AU2………………………………………………………………… 49

21 Expression intensity time course of AU1, 2 and 5 fit to the displacement changes of facial feature points based on the non-linear mapping [107]……….... 52

22 a. 1- and 2-dimensional scaling functions and wavelets [101]…………………... 54

	b. Dilation and translation of 1-dimensional basis (scaling and wavelet) functions [101]……	55
23	Automatic dense flow tracking for an image sequence. Out-of-plane motion (pitch) occurs at the bottom image. Dense flows are shown once for every 13 pixels……	61
24	Good tracking performance of using the wavelet-based dense flow for (a) furrow discontinuities at the forehead and chin regions, and (b) textureless regions with reflections at the forehead and cheek……	63
25	Tracking errors of the 2-level wavelet-based dense flow because of (a) large movements of brows or mouth, and (b) eye blinking also introduces motion error at brow regions……	64
26	Dense flow normalization using affine transformation: (a) includes both the rigid head motion in upward and leftward direction and non-rigid facial expression, and (b) eliminates the rigid head motion by using the affine transformation……	65
27	The mean (average) flow c is divided into horizontal flow c_h and vertical flow c_v for further processing by the principal component analysis (PCA)……	67
28	Each dense flow image is divided into horizontal and vertical flow images for the principal component analysis (PCA)……	72
29	a. Computation of eigenflow (eigenvector) number for the upper facial expressions: (a.1) is for the horizontal flow and (a.2) is for the vertical flow. The compression rate is 93:1 (932:10) and from which the recognition rate is 92% based on 45 training and 60 testing image sequences……	73
	b. Computation of eigenflow (eigenvector) number for the lower facial expressions: (b.1) is for the horizontal flow and (b.2) is for the vertical flow. The compression rate is 80:1 (1212:15) and from which the recognition rate is 92% based on 60 training and 90 testing image sequences……	74

30 Principal component analysis (PCA) for (a) horizontal flow weight vector and (b) vertical flow weight vector for the upper facial expressions (M' = 10)……. 75

31 Expression intensity matching by seeking the minimum distance between the weight vector of the testing frame and the weight vectors of all frames in a training sequence, whose expression intensity values are known. Each weight vector of the training image corresponds to a given expression intensity value.. 79

32 a. High gradient component (furrow) detection for the forehead and eye regions…………………………………………………………………….. 82

b. High gradient component (furrow) detection for the mouth, cheek, and chin regions…………………………………………………………………….. 83

c. High gradient component (furrow) detection for the chin region…………... 84

33 Permanent furrows or hair occlusion………………………………………... 85

34 The procedure of the high gradient component analysis in the spatio-temporal domain, which can reduce the effect of the permanent high gradient components (furrows) and hair occlusion for the upper facial expression…….. 88

35 (a) Original gray value images. (b) High gradient component (furrow) detection in the spatial domain. (c) High gradient component analysis in the spatio-temporal domain………………………………………………………... 89

36 a. High gradient component detection with different constant threshold values…………………………………………………………………………... 90

b. High gradient component detection with different constant threshold values…………………………………………………………………………... 91

37 a. Younger subjects, especially infants (Figure 37.a), show smoother furrowing than older ones (Figure 37.b), and initial expressions show weaker furrowing than that of peak expressions for each sequence……………………. 92

b. Younger subjects, especially infants (Figure 37.a), show smoother furrowing than older ones (Figure 37.b), and initial expressions show weaker furrowing than that of peak expressions for each sequence……………………. 93

38 a. Delete redundant high gradient components using morphological transformation including erosion and dilation processings……………………..…. 94
b. Delete the redundant high gradient components using the connected component labeling algorithm……………………………………………….…. 95
39 Horizontal line (furrow) detection using different sizes of detectors. (b) If the size of the detector is too small compared with the width and length of the line (furrow), then each line will be extracted to two lines. (c) It is necessary to adjust the size of the line detector to match the width and length of the line in order to obtain the correct result………………………………………..….. 97
40 Teeth can be extracted directly from the subtraction of the gray value image at the current frame to that at first frame for each image sequence whose absolute value is larger than a constant threshold……………………………………….. 99
41 Mean-Variance vector of the high gradient component analysis in the spatio-temporal domain for input to the Hidden Markov Model……………………… 100
42 Furrow expression intensity matching by measuring the minimum value (distance) of the sum of squared differences (SSD) between the mean-variance vector of the known training image and that of the testing image. Each mean-variance vector of the training image corresponds to a given expression intensity value…………………………………………………………….. 103
43 Vector quantization for encoding any vector sequence to a symbol sequence based on the codebook……………………………………………………….110
44 The construction (topology) of the Hidden Markov Model…………………….113
45 The tree structure of the computational complexity for direct evaluation of the output probability $P(O|\lambda)$ [105]………………………………………………….. 116
46 The Forward and Backward Procedures……………………………………..119
47 The tree structure of the computational complexity for the forward and backward procedures [79]………………………………………………………..120

48 A posterior probability variable $\gamma_t(i)$ which is the probability of being in state i at time t by given the HMM parameter set λ and the entire observable symbol sequence O………………………………………………………………... 126
49 The probability variable $\xi_t(i,j)$ which represents the probability of being in state i at time t, and state j at time t+1 given the observable symbol sequence O and the HMM parameter set λ……………………………………………………..…. 126
50 (a) 4 state ergodic HMM (b) 1st-order 4-state left-right (Bakis) HMM (c) 2nd-order 4-state left-right HMM………………………………………………... 129
51 A 1st-order 3-state Markov Model used to represent the observable symbol sequence………………………………………………………………….... 140
52 A 1st-order Hidden Markov Model can be used to represent the combination of all 1st-order Markov Models for facial expression AU12…………………... 143
53 A 1st-order Hidden Markov Model can be used to represent the combination of all 1st-order Markov Models for facial expression AU12…………………... 145
54 A 1st-order Hidden Markov Model can be used to represent the combination of all 1st-order Markov Models for facial expression AU15+17…………….… 146
55 A 1st-order Hidden Markov Model can be used to represent the combination of all 1st-order Markov Models for facial expression AU15+17…………….… 148
56 A 1st-order Hidden Markov Model can be used to represent the combination of all 1st-order Markov Models for facial expression AU17+23+24…………... 150
57 A 1st-order Hidden Markov Model can be used to represent the combination of all 1st-order Markov Models for facial expression AU6+12+25………….… 152
58 A 1st-order Hidden Markov Model can be used to represent the combination of all 1st-order Markov Models for facial expression AU6+12+25………….… 155
59 A 1st-order Hidden Markov Model can be used to represent the combination of all 1st-order Markov Models for facial expression AU6+12+25………….… 158
60 The 1st-order 2-, 3- and 4-state Hidden Markov Models can combine to be a 3rd-order 4-state Hidden Markov Model……………………………………….161

61 Experimental setup……………………………………………………………...165

62 Each facial expression begins from the beginning duration, continues through the apex duration, and ends at the ending duration. In our current work, we segmented each facial expression to include only the beginning and apex durations……………………………………………………………………….. 168

63 a. Standard AU1+4 expressions and manual misclassification of three AU1 expressions and one AU1+2 expression to AU1+4 expressions……………….. 169

b. Manual misclassification of three AU4 expressions to AU1+4 expressions. These mistakes are because of (1) Ω shape of furrows at the forehead, (2) confusing expression, and (3) asymmetric brow motion. The standard AU4 expression is shown in Figure 63.c (3)……………………………………….… 170

c. Confusions among AU12+25, AU20+25 (also in Figure 63.a and 63.b) and AU12+20+25……………………………………………………………………... 170

64 Three sets of extracted information as inputs to the recognition system using Hidden Markov Models…………………………………………………………... 172

65 The images at the same row have the same facial expressions, but different facial actions or expression intensities…………………………………………. 173

66 The training process for the Hidden Markov Model (an example for the lower facial expressions: AU12, AU6+12+25, AU20+25, AU9+17, AU17+23+24 and AU15+17)………………………………………………………………. 176

67 The recognition process for the Hidden Markov Model (an example for the lower facial expressions: AU12, AU6+12+25, AU20+25, AU9+17, AU17+23+24 and AU15+17)…………………………………………………….. 178

68 Different FACS AUs have the similar shape of furrows such as between AU12 and AU6+12+25, between AU6+12+25 and AU20+25, and between AU9+17 and AU17+23+24, since there are common facial muscle actions for both facial motions…………………………………………………………………………... 185

LIST OF TABLES

Table No. Page

1 Correspondence between facial expressions and elements of the Hidden Markov Model………………………………………………………………. 18

2 Comparison of modeling facial expressions with modeling speech using HMMs……………………………………………………………………….. 20

3 Action Units (AUs) in the Facial Action Coding System (FACS) [34]………….. 21

4 Sample symbol sequences for three upper facial expressions and six lower facial expressions under consideration……………………………………. 48

5 The dictionary for expression intensity estimation (image: 417 x 385 pixels).…. 50

6 Sample symbol sequences for three upper facial expressions and six lower facial expressions under consideration………………………………….… 76

7 Sample symbol sequences for three upper facial expressions and six lower facial expressions under consideration………………………………….… 101

8 Physical meaning of the Hidden Markov Model topology……………………... 163

9 Different Hidden Markov Models for 3 upper facial expressions and 6 lower facial expressions………………………………………………………………..174

10 The trained parameter set $\lambda = (\pi, A, B)$ of the 3rd-order 4-state Hidden Markov Model, whose topology is determined in Figures 58 and 60, for the lower facial expression AU6+12+25 using dense flow tracking method (codebook size M=16)……………………………………………………………………….. 177

11 The number of the training and testing image sequences (the average number of frames per image sequence is 20) and their corresponding recognition rates…………………………………………………………………………..180

12 Recognition results of the feature point tracking method. (The number given in each block is the number of testing image sequences.)……………………… 181

13 Recognition results of the dense flow tracking method. (The number given in each block is the number of testing image sequences.)…………………………..... 182

14 Recognition results of the motion furrow detection method. (The number given in each block is the number of testing image sequences.)………………....183

15 Summary of three different extraction and recognition methods for facial expressions………………………………………………………………….... 186