Stony Brook University

Analysis and Visualization of Facial Expression Subtlety Using Dynamic High Resolution Data

A Thesis Presented by

Lei Zhang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Computer Science

Stony Brook University

May 2008

Stony Brook University

The Graduate School

Lei Zhang

We, the thesis committee for the above candidate for the Master of Science degree, hereby recommend acceptance of this thesis.

Dimitris Samaras – Thesis Advisor

Associate Professor, Computer Science Department

Klaus Mueller - Chairperson of Defense

Associate Professor, Computer Science Department

Xianfeng David Gu

Assistant Professor, Computer Science Department

This thesis is accepted by the Graduate School

Lawrence Martin

Dean of the Graduate School

Abstract of the Thesis

Analysis and Visualization of Facial Expression Subtlety

Using Dynamic High Resolution Data

By

Lei Zhang

Master of Science

In

Computer Science

Stony Brook University

2008

Facial expression is an important visual channel for the display of human emotions. Automatic expression analysis and understanding is in the core of natural Human-Computer Interface (HCI) design. Although expression was extensively studied in behavior science since early 1970s, automatic analysis of facial expression did not start until 1990s. In this thesis, we compared different expression databases used in computer vision field and found out the lack of a publicly available dynamic 3D expression database and a systematic evaluation platform of expression analysis algorithms. We reviewed various automatic expression analysis systems and concluded that most of current systems analyze deliberated expression captured in controlled environment. However, there is significant difference between deliberated and spontaneous expressions in terms of timing, symmetry and intensity. In practice, challenges like illumination change, head motion, partial occlusion and speeches would complicate the analysis of spontaneous expression. We solve the problem of expression analysis by utilizing dynamic high resolution data. Dense motion flows are computed from facial expression sequences and visualized in the 2D space using Linear Integral Convolution. Experimental results show clear motion patterns in our facial expression data which are verified by observations of researchers.

Table of Contents

List of Figures ..... **v**

List of Tables ..... **vii**

1 Introduction **1**

2 Expression Data Collection **3**

2.1 Expression Datasets in Face Recognition Database 5
2.2 2D Facial Expression Databases 5
2.3 3D Facial Expression Databases 10

3 Automatic Analysis of Facial Expression **12**

3.1 Representation of Facial Expression 13
3.2 Spatial Analysis of Facial Expression 15
3.3 Spatial-temporal Analysis of Facial Expression 15
3.4 Analysis-by-Synthesis 30
3.5 Evaluation of Expression Analysis Algorithms 30

4 Analysis and Visualization of 3D High Resolution Expression Data **32**

4.1 Data Preprocessing 32
4.2 Image Data for Flow Computation 33
4.3 Optical Flow Computation and Visualization 33
4.4 Flow Result Verification 34
4.5 Failed Cases and Why 35

5 Expression Analysis by Using High Speed Video Camera **37**

5.1 Data Set 37
5.2 Flow Computation 38
5.3 Flow Visualization 38
5.4 Statistical Analysis of Flow Data 40
5.5 Comparison of Flow Results from Different People 41

6 Summary **48**

References **50**

List of Figures

Figure 1.1 Historical view of facial expression analysis 1
Figure 2.1 Samples from JAFFE database (adapted from [Lyons etc.99])..... 7
Figure 2.2 A sample image from Ekman-Hager expression database. The image shows upper facial muscles corresponding to action units 1,2,4,6 and 7. [Ekman etc. 99] 7
Figure 2.3 the interface for the UA-UIUC authentic expression analysis system. One sample image of the database is shown on the right with a tracked wireframe overlapped on the face of the subject. [Sebe and Lew etc. 04]..... 9
Figure 2.4 Data collection set up. a. The frontal camera was mounted in a bookshelf above the interrogator's head. b. Two side view cameras were wall-mounted. A fourth camera was mounted under the interrogator's chair. c. Interrogators were retired members of the police and FBI. [Frank and Ekman 97] 10
Figure 2.5 Sample models from BU-3DFE databases. (a). our levels of facial expressions from low to high. Expression models show the cropped face region and the entire facial head. (b). even expressions male (neutral, angry, disgust, fear, happiness, sad, and surprise), with face images and facial models [Yin etc. 06] 11
Figure 3.1 The cues for facial expression as suggested by Bassili [Black and Yacoob 97] 16
Figure 3.2 Three phases of a facial expression[Yacoob and Davis 96] 17
Figure 3.3 a-d the upper left quadrant shows the intensity image, the upper right quadrant show the gradient image, the rectangle in between displays the classification of facial expression, the lower left quadrant show the optical flow field, the rectangles around the face regions of interest and the mapping of colors into directions, and the lower right quadrant shows the mid-level descriptions that were computed. [Yacoob and Davis 96]..... 18
Figure 3.4 The figure illustrates the motion captured by the various parameters used to represent the motion of the regions. The solid lines indicate the deformed image region and the "-" and "+" indicate the sign of the quantity [Black and Yacoob 97]. 18
Figure 3.5 Additional parameters for planar motion and curvature [Black and Yacoob 97]. 19
Figure 3.6 Determining of expressions from video sequences (a) and (b) show expressions of smile and surprise, (c) and (d) show a 3D model with surprise and smile expressions, and (e) and (f) show the spatial-temporal motion energy representation of facial motion for these expressions. [Essa and Pentland 97] 21
Figure 3.7 Structure of Potential Net and nodal deformation [Matsuno 95] 21
Figure 3.8 19 FFPs. [Wang etc. 98] 22
Figure 3.9 The FFP tracking system [Wang etc. 98]..... 22
Figure 3.10 Multi-state Facial Component Models of a Frontal Face [Tian et al. 01] 24

Figure 3.11 Multilevel HMM architecture for automatic segmentation and recognition of emotion [Cohen et al. 03] 25
Figure 3.12 BN for vigilance detection [Gu and Ji 04] 26
Figure 3.13 (a) A posed image sequence performing occluded facial expression. (b) The result from our facial expression model. [Gu and Ji 04] 27
Figure 3.14 Illustration of a 3D expression manifold. The reference center is defined by the neutral face. Image sequences from three different expressions are shown. The further a point is away from the reference point, the higher is the intensity of that expression. [Chang 2005] 28
Figure 3.15 (a) The prior BN for AU modeling before learning; (b) The learnt BN from the training data [Tong and Ji 06] 29
Figure 4.1 One sample frame 32
Figure 4.2 Frame No. 294 33
Figure 4.3 Frame No. 298 33
Figure 4.4 Flow result: 294->298 34
Figure 4.5 Comparison between ground truth data and optical flow result 35
Figure 4.6 A failed case when head motion is introduced in the expression data 36
Figure 5.1 One sample frame from a surprise expression 38
Figure 5.2 LIC visualization of the flow results 39
Figure 5.3 Snapshot of IBFV results 40
Figure 5.4 Deformation statistics around eyes and mouth 41
Figure 5.5 Comparison of different flow results 47

List of Tables

Table 2.1 Review of facial expression databases for automatic expression analysis 4
Table 2.2 CMU-Pittsburgh AU-Coded Facial Expression Database [Cohn etc, 00] 6
Table 3.1 Single Action Units (AU) in FACS [Ekman and Friesen 78] 14
Table 3.2 More grossly defined AUs in the FACS [Ekman and Friesen 78] 14

1 Introduction

Facial expression is an important channel for effective communication between humans. In people's daily conversations, speakers read the internal emotional state of listeners from their expressions and feedback appropriately to the listeners. For an ideal Human-Computer Interaction (HCI) system, we expect the computer will be able to recognize different facial expressions from users and give suggestions to the users intelligently. For example, a driver fatigue detection system should be able to detect different vigilance levels of the driver by tracking and analyze his/her facial actions and alert the driver when an inattention or yawning event is detected [Gu and Ji 04].

During the past decades, lots of researchers in behavioral science try to characterize and categorize human expressions and set up the relationship between expression characteristics and internal emotional states of humans. Some interesting applications of this research include deception detection [Ekman and Friesen et al. 88], depression analysis [Heller and Haynal 94], Nonverbal expressions of psychiatric patients [Ellgring 86] and so on. Although the research for human expression analysis in behavior science has been initiated since 1970s, automatic expression analysis using computers didn't start until 1990s when researchers in computer vision and graphics area began to track faces from digitalized images, extract features from the images and categorize them into different expression classes. An overview of these analysis techniques are shown in Figure 1.1

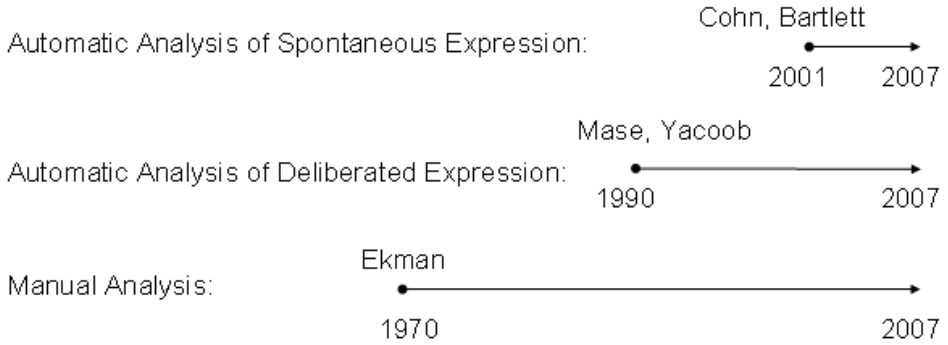


Figure 1.1 Historical view of facial expression analysis

Up to now, computers prove to be able to recognize deliberated expression with

accuracy of more than 90 percent in controlled environment by using spatial-temporal analysis techniques. However, how to recognize spontaneous expression in an uncontrolled environment remains an open question.

The work on automatic expression analysis could be divided into two categories. The first category contains systems using only spatial information extracted from input images. These images usually are "mug-shot" views of an expression at its peak. By comparing features extracted from current input image with those from a reference neutral face image, the algorithms compute the feature displacement and recognize the expression based on some predefined heuristics. While these algorithms may work in controlled environment, they don't take dynamic nature of human expression into account. The timing and temporal evolution of facial expression has been proven to be important to differentiate deliberated and spontaneous expressions. [Schmidt and Cohn et al. 06] Also, challenging issues rising in spontaneous expression analysis such as partial occlusion due to head motion, ambiguity and low density of expressions would easily fail these systems. The second category contains systems which utilize both spatial and temporal information from input image sequences. By explicitly modeling the temporal behavior of expression, these systems could completely represent the whole process of expression evolution thus are usually more robust than previous systems.

A common disadvantage for most of the above systems is that they all work in relatively controlled environment in which out-of-plane head motion of subjects is very limited and no partial occlusion is allowed in the input image sequences. Spontaneous human expressions, however, are always accompanied by different head rotation, speeches and other body movement. Low intensity, hair style, glasses and beard could complicate the analysis of facial expression as well. In a word, the adaptation of these systems to spontaneous expression data is a non-trivial problem. On the other hand, behavioral scientists found that there is significant difference between deliberated and spontaneous expressions [Schmidt and Cohn et al. 06] which turns out to be very important to analyze the internal emotional state of humans for psychological applications such as deception detection. This asks researchers in computer vision community for more robust and precise facial expression analysis systems which could track 3D motion of facial features with different density, handle missing features due to partial occlusion and model dynamics of facial expressions. Motivated by these requirements, groups at CMU/Pittsburg, UCSD and RPI [Cohn and Kanade etc 06, Barlett and Frank etc. 05, 06, Zhang and Ji 05] began to investigate spontaneous expression analysis problem and partially solved the problem from different perspectives. A detailed description could be found in Section 3 of this paper. To the best of authors' knowledge, the automatic analysis of spontaneous expression is a new born [Bartlett etc. 01]. Lots of issues (part of them are listed above) are still great challenges for computer vision researchers. In this paper, we will review some of these techniques developed from

1990s, summarize pros and cons of them and cast some light into future's research on spontaneous facial expression analysis.

The rest of this paper is organized as follows: In Section 2, we first talk about issues in expression data collection and introduce facial expression databases which are widely used in computer vision research community before we go through different expression analysis techniques in Section 3. In Section 4, we analyze and visualize facial expression subtlety using dynamic high resolution 3D data. Following that in Section 5, we try to prove that with a high speed camera, we could capture and analyze motion patterns in higher temporal resolution of facial expression. Section 6 will conclude the paper and discuss the challenges arose in spontaneous facial expression analysis and cast some lights to the future work.

2 Expression Data Collection

Comparing to generic object motion in natural scene, facial motion is more complex. The motion of facial features is combination of rigid head motion and non-rigid facial deformation. For expression data captured using a normal video camera, the 3D head motion is usually unknown. Also, the muscle activities which control non-rigid facial deformation are unknown either. To simplify the automatic expression recognition problem, some of currently available expression databases record facial expressions in front view under controlled lab environment. While this makes facial detection and tracking much easier, it imposes great limitation on the richness of captured expressions.

According to [Picard 01], there are five important factors which suggest a natural setup for gathering genuine spontaneous facial expression data:

1. *Subject-elicited versus event-elicited*: Does subject purposefully elicit emotion or is it elicited by a stimulus or situation outside the subject's efforts?
2. *Lab setting versus real-world*: Is subject in a lab or in a special room that is not their usual environment?
3. *Expression versus feeling*: Is the emphasis on external expression or on internal feeling?
4. *Open-recording versus hidden-recording*: Does subject know that anything is being recorded?
5. *Emotion-purpose versus other-purpose*: Does subject know that the experiment is about emotion?

Although some of these issues are ad hoc to the analysis of affective physiological state, we could still find these factors important to spontaneous expression data collection. Ideally subjects should sit in some location which is familiar to them (real-world). Their expressions should be elicited by an external event of circumstances (event-elicited). And they should be unaware of capturing (hidden-recording) and experimental purpose of expression data collection (other-purpose). Unfortunately, it's usually hard to set up this ideal capturing setting to collect data for our experiments. How to build a public spontaneous expression database which satisfies most of these five suggestions is also an open question for future research on spontaneous expression.

Table 2.1 shows some well-known expression databases reported from 1990s. Some of them are subsets of face recognition databases and contain only static images or 3D models which are "mug-shot" views of an expression sequence at its apex. The others contain video sequences for each subject with different expressions thus could be used for dynamics modeling of expressions.

Table 2.1 Review of facial expression databases for automatic expression analysis

Database	Static images/models	Dynamic sequences
2D	CMU-PIE [Sim and Baker 03], Yale [Belhumeur etc 97], AR [Martinez etc. 98], ORL [Samaria and Harter 94], FERET [Phillips etc. 00], FRGC [Phillips etc. 05], Lots of others [Zhao etc. 03]	Cohn-Kanade [Cohn etc, 00], JAFFE [Lyons etc.99], Ekman-Hager [Ekman etc. 99], RPI ISL [Zhang and Ji 05], UA-UIUC [Sebe and Lew etc. 04] RU-FACS-I [Barlett and Frank 06], Frank-Ekman [Frank, Ekman 97],
3D	BU-3DFE [Yin etc. 06]	None

2.1 Expression Datasets in Face Recognition Database

Some of publicly available face recognition databases contain expression change in subsets of their face images. CMU-PIE database consists of more than 40,000 images of 68 subjects. Each subject has 640*486 images from 13 poses under 43 different illumination conditions. Four expressions (neutral, smile, blink and talk) are recorded for each subject. Yale database contains 165 320*243 images from 15 subjects with 6 different expressions (neutral, happiness, sadness, sleepiness, surprise, and wink) under 3 different lighting conditions (center-light, left-light, and right-light). AR face database consists of 4000 768*576 images from 126 subjects (70 male and 56 female) with 4 different facial expressions (neutral, smile, anger and scream) under 3 illumination conditions (left, right and all lights on). ORL database was developed by AT&T Lab, Cambridge. It captures 400 92*112 images from 40 subjects with 3 facial expressions (neutral, smiling, closed eyes). Both FERET and FRGC databases are primarily used for face recognition instead of facial expression analysis although they do capture limited expression variation in the database. [Zhao etc. 03] gives an in-depth survey of these databases and their applications in face recognition context.

2.2 2D Facial Expression Databases

Different from databases described in Section 2.1, the databases introduced here are primarily used for facial expression analysis. Part of them record spontaneous expression for psychological experiments while the others capture posed expression for pattern recognition purpose. Cohn-Kanade AU-coded database [Cohn etc, 00] is the most popular one among these databases. It includes video sequences from 210 adults between 18 and 50 years old. 69 % of them are female, 31% male, 81% Euro-American, 13% Afro-American, and 6% other groups. Facial behaviors of these subjects were captured using two Panasonic WV3230 cameras, each connected to a Panasonic AG-7500 video recorder with a Horita synchronized time-code generator. One camera was right in front of the subjects and the other one was 30 degrees to the subject's right. Each subject was instructed by an experimenter to perform a series of 23 facial displays. All the image sequences were then digitalized into 640 by 480 or 490 8-bit grayscale images. A subset of these image sequences are released to research community. The subjects in this subset are 100 university students with the age between 18 and 30. Sixty-five percent were female, 15 percent were African-American, and three percent were Asian or Latino. Only front view video from the subjects is available in this release. Each expression sequence begins from neutral state and ends at the apex of the expression. The last frame of the sequence was manually coded by a certificated FACS coder and 17% of these coding are verified by a second certificated FACS coder. A summary of Cohn-Kanade AU-coded

database is shown in Table 2.2.

Table 2.2 CMU-Pittsburgh AU-Coded Facial Expression Database [Cohn etc, 00]

Subjects	
Number of subjects	210
Age	18-50 years
Women	69%
Men	31%
Euro-American	81%
Afro-American	13%
Other	6%
Digitized sequences	
Number of subjects	182
Resolution	640x490 for grayscale 640x480 for 24-bit color
Frontal view	2105
30-degree view	Video-type only
Sequence duration	9-60 frames/sequence

The Japanese Female Facial Expression (JAFFE) Database [Lyons etc.99] was collected in Psychology Department at Kyushu University. It consists of 213 static images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female subjects. Each expression is rated on one of six prototypic expression categories (Happiness, Sadness, Surprise, Fear, Angle and Disgust) by 60 Japanese people. Figure 2.1 illustrates some sample images from JAFFE database.

In [Ekman etc. 99], Ekman, Hager, etc. constructed an expression database of more than 1100 sequences containing over 150 facial actions or action combinations from 24 subjects. The subjects were instructed by a FACS expert to perform different expressions. The images were then coded by three certificated FACS coders to provide ground truth for evaluation of automatic AU recognition algorithms. A sample is shown in Figure 2.2

with underlying muscles overlapped on the image.

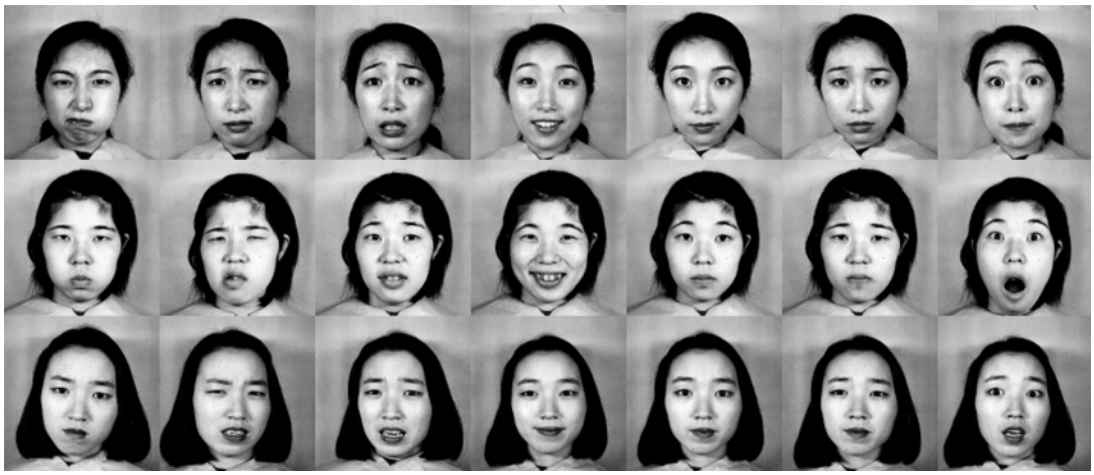


Figure 2.1 Samples from JAFFE database (adapted from [Lyons etc.99])

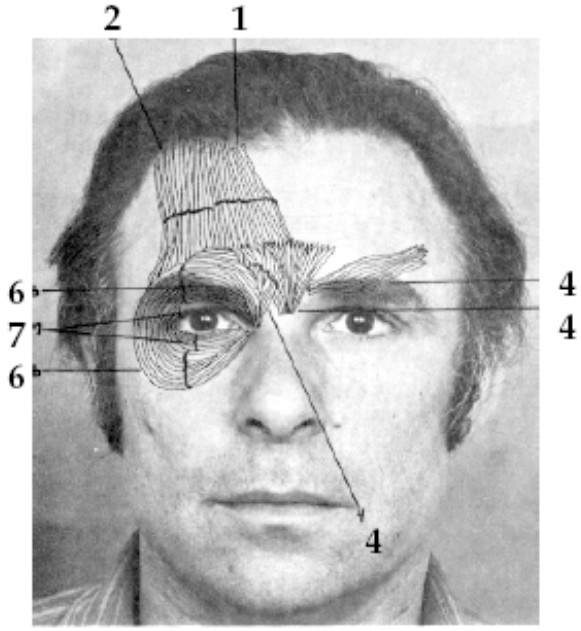


Figure 2.2 A sample image from Ekman-Hager expression database. The image shows upper facial muscles corresponding to action units 1,2,4,6 and 7. [Ekman etc. 99]

The Rensselear Polytechnic Institute Intelligent Systems Lab (RPI-ISL) facial

expression database consists of two subsets [Zhang and Ji 05]. One subset contains 42 color video sequences from 10 subjects with various expressions but fixed frontal-view head pose at 30 fps. The resolution of images is 320 by 240. Each expression begins from a neutral state, goes to the apex and comes back to the neutral state. Each frame of expression video sequences is coded by FACS coders. Manually labeled AUs together with the identifications of the subjects provide ground truth for algorithm verifications. The second subset is primarily used for facial action unit recognition under head movements. It contains 40 color image sequences from 8 subjects. The expression data was captured by using the same experimental setting as that used by the first subset of this database. But in this subset natural head movements are recorded from five different directions. In addition to the labeled AUs and identity information provided in the first dataset, this dataset provides coordinates of pupils and 34 facial feature points for selected frames as ground truth for face tracking and expression analysis.

Up to this point, we reviewed several databases containing only posed expression data. i.e. the subjects' expressions are elicited by an experimenter or expert thus not their natural response to the stimuli from the environment. Some researchers in behavior science and computer vision area try to collect spontaneous expression data with more natural capturing settings. Below we will introduce these databases and describe their solutions to the challenges presented at the beginning of this section.

Developed by Sebe etc, UA-UIUC [Sebe and Lew etc. 04] database contains spontaneous expression data from 28 students within the computer science department. Students sat in a video kiosk and watched a fragment of movie trailers. Their response to the video trailers is recorded by a hidden video camera. After the test the experimenter interviewed with students and found out the true emotional state corresponding to their expressions in the video footage of hidden cameras. A snapshot of their system interface with a sample face image is shown in Figure 2.3.

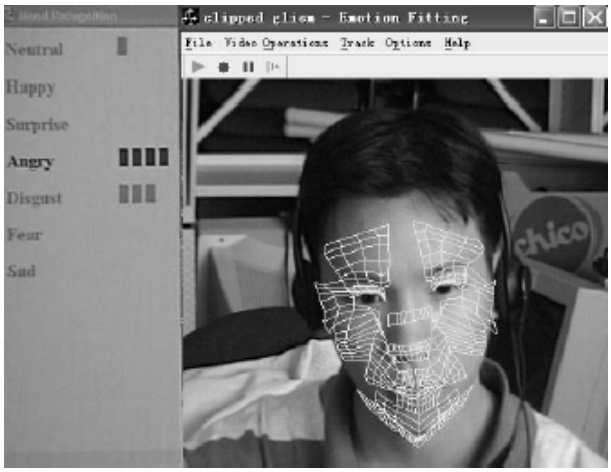


Figure 2.3 the interface for the UA-UIUC authentic expression analysis system. One sample image of the database is shown on the right with a tracked wireframe overlapped on the face of the subject. [Sebe and Lew etc. 04]

Bartlett etc. at UCSD collaborated with Mark Frank from Rutgers University (now is at SUNY Buffalo) and created a comprehensive expression database – RU-FACS-1 [Barlett and Frank 06]. It contains spontaneous expressions from 100 subjects with multiple views. The average length of expression is 2.5 minutes. Subjects participated in a 'false opinion' paradigm in which they were asked their opinions about a social or political issue. If the subject takes the opposite opinion on the issue and successfully convinces the interviewer he/she is telling the truth, he/she will receive $50. On the other hand, however, if they are not believed regardless whether they are lying or telling the truth, they will not be able to get any money and instead have to fill out a long boring questionnaire. This database is challenging due to speech-related expression and distinct out-of-plane head motion. The expression data was labeled by two certificated FACS coders to provide ground truth for algorithm evaluation. Figure 2.4 shows the data-collection set up and a sample image within the database.

Figure 2.4 Data collection set up. a. The frontal camera was mounted in a bookshelf above the interrogator's head. b. Two side view cameras were wall-mounted. A fourth camera was mounted under the interrogator's chair. c. Interrogators were retired members of the police and FBI. [Frank and Ekman 97]

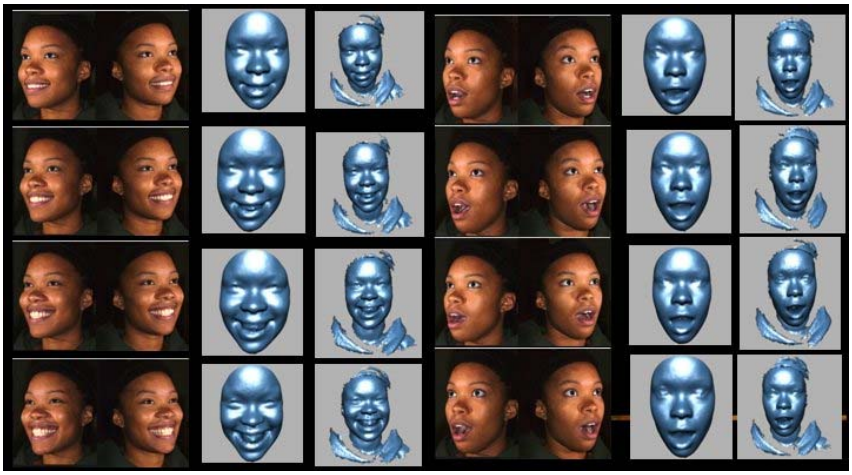
Another reported spontaneous expression database is presented in [Frank and Ekman 97]. In Frank-Ekman database, 20 young adult men (7 Euro-American, 2 African-American and 1 Asian) were asked if they had stolen a large amount of money. If they could successfully convince the interviewer they didn't steal the money while in fact they did, they will receive as much as $50 otherwise they would expect serve punishment. Totally 12 subjects stole $50 and 8 told the truth. The facial expressions of subjects were video-recorded using an S-Video camera and then digitalized into 640 by 480 images with 16 bit color intensity. A certified FACS coder manually coded the video from the first 10 subjects. Challenging issues like non-frontal pose and out-of-plane head motion are common. Some of subjects wore glasses in the experiments.

2.3 3D Facial Expression Databases

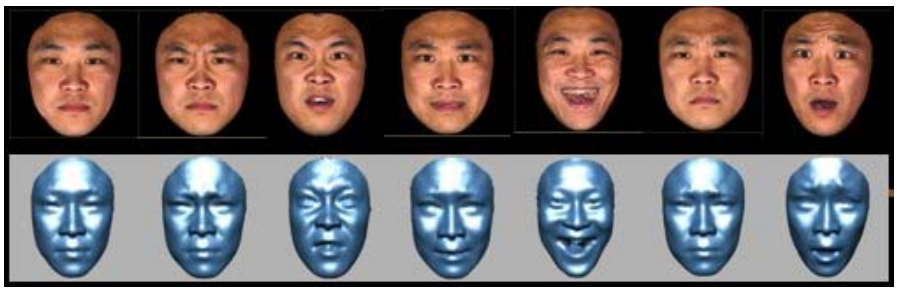
To overcome the limitation of 2D image data, some researchers try to capture expressions in 3D [Zhang etc. 04, Chai etc. 03]. Most of these 3D databases are mainly used for face expression modeling and animation in Graphics. It's not clear if they contain enough features for the use of expression analysis. But it's an interesting direction to go to find out what features are essentially important for facial expression understanding.

There is only one static 3D expression database [Yin etc. 06] reported for expression analysis recently - BU-3DFE. This database records 2500 static 3D face models from 100 subjects (56% female, 44% male) with a variety of ethnic/race background. The age of these subjects range from 18 to 70 years old. Each subject was asked to perform six basic expressions with four levels of intensity. Together with neutral face, there are 25 scans for

each subject in the database. For each geometric shape model, there are two corresponding texture images at $\pm 45^{\circ}$ viewpoints. Figure 2.5(a) shows four levels of expression intensity and corresponding two view texture and geometric models. (b) demonstrates 7 different basic expressions and their corresponding face models.



(a)



(b)

Figure 2.5 Sample models from BU-3DFE databases. (a). our levels of facial expressions from low to high. Expression models show the cropped face region and the entire facial head. (b). even expressions male (neutral, angry, disgust, fear, happiness, sad, and surprise), with face images and facial models [Yin etc. 06]

To the best of author's knowledge, there is no dynamic 3D facial expression database reported in computer vision and psychology literatures so far. This is definitely a good starting point to solve facial expression analysis problem in 3D which the author think is the future of robust automatic expression analysis.

In this section, we surveyed most of facial expression databases reported in

literatures. As pointed out by [Cohen etc. 03], a standard publicly available benchmark database is important for the progress of automatic expression analysis. At this point, Cohn-Kanade database seems a good candidate for posed expression analysis in 2D. For spontaneous expression analysis, however, there is no widely accepted database which could be used to test and compare different expression analysis systems. RU-FACS-1 is promising in the sense that it contains spontaneous expressions from large amount of subjects. 3D expression database collection is almost an untouched area due to its high cost and unavailability for most of expression researchers. But for an automatic expression analysis to work robustly in practice, 3D shape information of faces will be necessary to track head motion and facial deformation in 3D space instead of 2D plane. In addition to publicly available databases, evaluation techniques of expression analysis systems are also very important to the development of future systems. We will investigate these problems in the near future.

3 Automatic Analysis of Facial Expression

In section 2, we introduce different databases for the analysis of facial expressions. In this section, we will survey variously automatic expression analysis techniques. We don't attempt to exhaustively elaborate all the work from 1990s to now. Instead we selectively discuss some of them which demonstrate typical characteristics of one class of similar algorithms. Hopefully the discussion here would be helpful for readers to build up a whole image of automatic expression analysis and cast some light into future research in this area.

In section 3.1, we will introduce several well known forms of facial expression representation including Facial Action Coding System (FACS) [Ekman and Friesen 78] and universal emotion categories [Ekman 94]. Based on the focus of expression analysis, the systems presented in computer vision literatures could be classified into two groups: One contains systems using only static cues of facial expression and performing spatial analysis on static images. The other consists of systems utilizing both static and dynamic cues of expressions and performing both spatial and temporal analysis on an input video sequence. These techniques will be discussed in section 3.2 and 3.2 respectively. Different from the works in computer vision area in the above, graphics researchers try to develop systems which could create realistic facial expression animation. A byproduct of these systems is a parametric model of face which could be used to analyze facial expression in an analysis-by-synthesis fashion. We will discuss these techniques in section 3.4. As we mentioned in Section 2, the evaluation of expression analysis algorithms plays an essential role to the development of future expression analysis

systems. We will talk about this issue at the end of this section.

3.1 Representation of Facial Expression

[Ekman 94] claimed that there are six basic emotions: *happiness, sadness, surprise, fear, anger* and *disgust* which are universal and cultural independent. Although there are many questions about this study in the past years [Russell 94, Ekman 94], most of automatic facial expression systems developed by computer vision researchers perform an emotional classification based on Ekman's six categories of facial expression.

In practice, however, richness of facial expressions could be far more than these six universal expressions. i.e. the six emotions are only a very small set of human expression space. These basic emotional categories are not enough to represent the richness and complexity of motions in different face regions. In 1978, Ekman [Ekman and Friesen 78] et al. proposed the Facial Action Coding System (FACS) as a measurement tool for expression analysis. The FACS is a comprehensive and anatomically based system for measuring all visually discernible facial movement in terms of 44 unique Action Units (AUs). Table 3.1 illustrates single AUs and muscles involved in these AUs. And Table 3.2 shows other grossly defined AUs. The FACS allows for coding of expression intensity on a 5-point scale. FACS is widely used in behavior science for physiological analysis of facial expression [Ekman and Rosenberg 97]. However, it needs a certificated FACS coder manually coding the face images in frame-by-frame or slow-motion viewing, which is very labor-intensive and time-consuming. So experts in computer vision collaborate with psychologists to automate this coding procedure [Barlett and Frank etc. 06, Lien et al. 97, Tian et al. 2001]. But a fully automatic system which is robust and efficient for spontaneous expression analysis is still not available at this time due to richness and complexity of facial motion and limited accuracy of facial tracking and analysis algorithms.

Table 3.1 Single Action Units (AU) in FACS [Ekman and Friesen 78]

AU number	Descriptor	Muscular Basis
1.	Inner Brow Raiser	Frontalis, Pars Medialis
2.	Outer Brow Raiser	Frontalis, Pars Lateralis
4.	Brow Lowerer	Depressor Glabellae, Depressor Supercilli; Corrugator
5.	Upper Lid Raiser	Levator Palpebrae Superioris
6.	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7.	Lid Tightener	Orbicularis Oculi, Pars Palebralis
9.	Nose Wrinkler	Levator Labii Superioris, Alaeque Nasi
10.	Upper Lip Raiser	Levator Labii Superioris, Caput Infraorbitalis
11.	Nasolabial Fold Deepener	Zygomatic Minor
12.	Lip Corner Puller	Zygomatic Major
13.	Cheek Puffer	Caninus
14.	Dimpler	Buccinator
15.	Lip Corner Depressor	Triangularis
16.	Lower Lip Depressor	Depressor Labii
17.	Chin Raiser	Mentalis
18.	Lip Puckerer	Incisivii Labii Superioris; Incisivii Labii Inferioris
20.	Lip Stretcher	Risorius
22.	LipFunneler	Orbicularis Oris
23.	Lip Tightener	Orbicularis Oris
24.	Lip Pressor	Orbicularis Oris
25.	Lips Part	Depressor Labii, or Relaxation of Mentalis or Orbicularis Oris
26.	Jaw Drop	Masetter; Temporal and Internal Pterygoid Relaxed
27.	Mouth Stretch	Pterygoids; Digastric
28.	Lip Suck	Orbicularis Oris

Table 3.2 More grossly defined AUs in the FACS [Ekman and Friesen 78]

AU number	FACS name
19.	Tongue out
21.	Neck Tightener
29.	Jaw Thrust
30.	Jaw Sideways
31.	Jaw Clencher
32.	Lip Bite
33.	Cheek Blow
34.	Cheek Puff
35.	Cheek Suck
36.	Tongue Bulge
37	Lip Wipe
38.	Nostril Dilator
39.	Nostril Compressor
41.	Lid Droop
42.	Slit
43.	Eyes Closed
44.	Squint
45.	Blink
46.	Wink

Alternatively, there are several other facial measurement mechanisms used in behavior science such as maximally discriminative facial movement coding system

(MAX) [Izard 79] and facial electromyography (EMG). Different from FACS, Izard's MAX is theoretically derived and codes only facial configurations that Izard theorized correspond to universally recognized facial expressions of emotion [Ekman and Rosenberg 97]. For those expressions which are not well studied in theory, the MAX could fail to capture the relevant facial behavior. Facial EMG measures electrical potentials from facial muscles to infer their activities. As pointed out in [Ekman and Rosenberg 97] and [Picard etc. 01], the EMG is an obtrusive measuring method. When subjects notice their faces are being measured, their responses will be different and this may potentially interfere the facial behavior under investigation. Another disadvantage of EMG is that the measurement of one facial muscle is interfering with measurements from others. This may cause confusing observations of muscle activities. Because of this, EMG is usually used in physiology combined with other sensors.

3.2 Spatial Analysis of Facial Expression

Since 1990s, different approaches are presented for facial expression analysis from still images. Early works include [Cottrell etc. 91], [Kearney and McKenzie 93], [Hara etc. 92], [Matsuno etc. 93], [Rahardja etc. 91], [Ushida etc. 93] and [Vanger etc. 95]. Works in the late 1990s include [Edwards 98], [Hara etc. 97], [Hong 98], [Huang 97], [Lyons 99], [Padget 96], [Pantic 00a], [Yoneyama 97], [Zhang and Lyons etc. 98] and [Zhao and Kearney 96]. Since spatial analysis from static input images is not the focus of this paper, we will not elaborate all the techniques listed above. A detailed review of these techniques could be found in [Pantic and L. Rothkrantz 00b].

3.3 Spatial-temporal Analysis of Facial Expression

Although the above systems could correctly recognize facial expression or Action Unit (AU) by extracting spatial facial features from static images, they share a common disadvantage of disregarding dynamic nature of facial expression. In [Bassili 79], Bassili claimed that motion in a face image would allow people to identify human emotion even with minimal information about the spatial arrangement of features. In his experiments, as shown in Figure 3.1, subjects were asked to recognize emotion from image sequences in which only white dots on the dark surface of the face displaying the emotion are visible. Without knowing the texture and salient facial features, subjects still recognized all the expressions at above chance levels while they could only recognize happiness and sadness from static images. This result suggests that utilizing dynamic cues from image sequences will improve the performance of facial expression analysis systems.

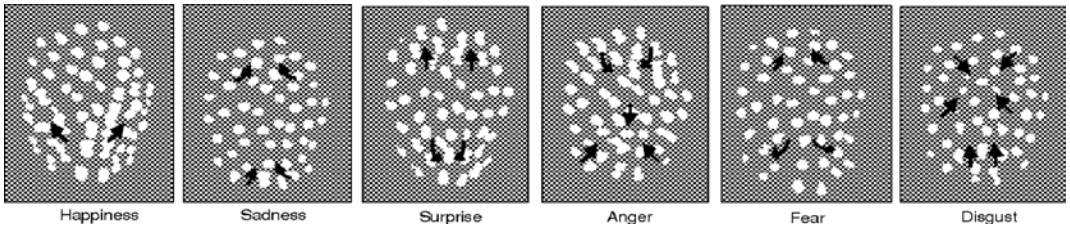


Figure 3.1 The cues for facial expression as suggested by Bassili [Black and Yacoob 97]

Early works capitalizing temporal information from image sequences include [Mase 91], [Moses etc. 95], [Rosenblum etc. 94] and [Yacoob and Davis 94, 96]. Mase is the first one to use image processing techniques (optical flow estimation) to recognize facial expressions. [Moses etc. 95] presented a system which fits a quadratic B-spline to a valley contour between two lips and tracks the contour using Kalman filter. The system works in real time and could recognize 5 different mouth shapes with a 100% recognition rate. The number of tested subjects was not reported in the paper though. [Yacoob and Davis 94, 96] computes optical flow from input image sequences and then perform spatial-temporal analysis on the flow. In [Rosenblum etc. 94], a region tracker was created based on the results from optical flow estimation. Tracked feature points were then fed into a Radial Basis Function Neutral Network to recognize basic expressions. An average recognition rate of 80% was reported.

[Yacoob and Davis 96] derive translation and scaling of a rectangle centered at some facial features from optical flow in the rectangle. Then a middle level linguistics description is used to infer the motion of facial component based on the motion of rectangles. After that a rule based system is used to identify the onset, apex and offset phases of expressions (see Figure3.2). Figure 3.3 shows the detection of the beginning of a ‘fear’ expression from the paper. As pointed out in [Zhang and Ji 05], optical flow estimation is easily disturbed by the change of illumination and non-rigid motion. Also it’s sensitive the alignment of images and motion discontinuity. So optical flow based systems require bounded facial motion, constant lighting and plenty features in local facial areas. Out-of-plane head motion and transient facial motion (wrinkles) will easily fail the systems.

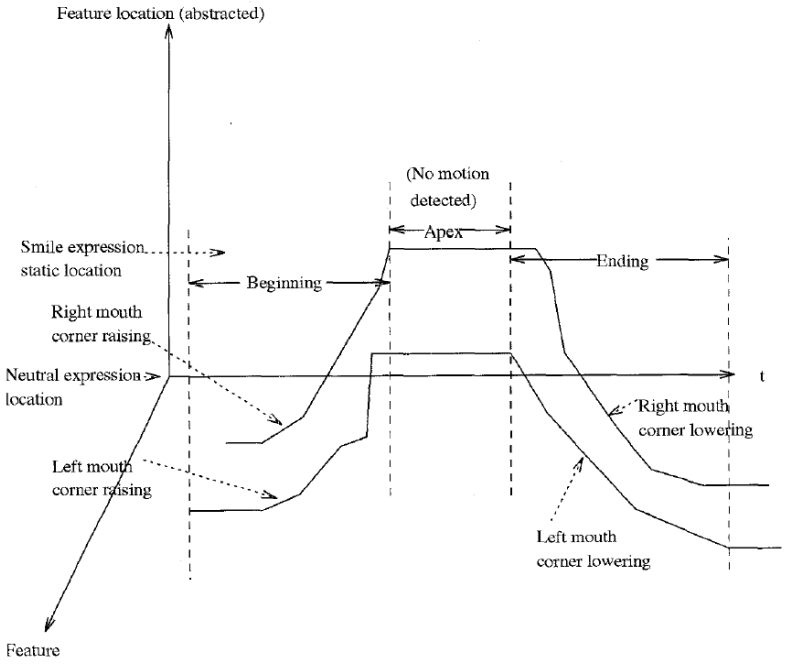


Figure 3.2 Three phases of a facial expression[Yacoob and Davis 96]

In the late 1990s, a bunch of systems [Black and Yacoob 97, Cohn 98, Essa 97, Matsuno 95, Kimura and Yachida 97, Wang 98] using different facial deformation models were presented. A good survey of these works could be found in [Pantic and L. Rothkrantz 00b].

[Black and Yacoob 97] presented a few local parameterized models to describe facial motion. A planar approximation was used to represent rigid head motion in a variety of situations. An affine-plus-curvature parametric model is used to describe non-rigid facial motion in stabilized facial images. Figure 3.4 and 3.5 demonstrate the affine and curvature model used in the paper. A robust regression algorithm based on brightness constancy assumption is employed to estimate facial motion parameters. Once the motion parameters are known, a mid-level and high-level rule based system is used to classify the facial expression into six basic categories. Similar to [Yacoob and Davis 96], each expression was divided into three temporal segments as shown in Figure 3.2.

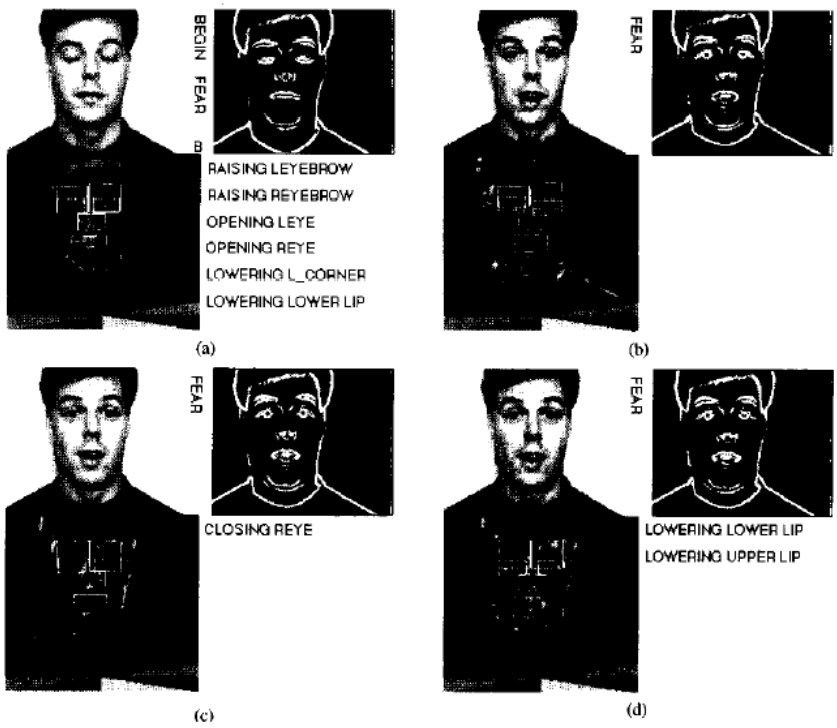


Figure 3.3 a-d the upper left quadrant shows the intensity image, the upper right quadrant show the gradient image, the rectangle in between displays the classification of facial expression, the lower left quadrant show the optical flow field, the rectangles around the face regions of interest and the mapping of colors into directions, and the lower right quadrant shows the mid-level descriptions that were computed. [Yacoob and Davis 96]

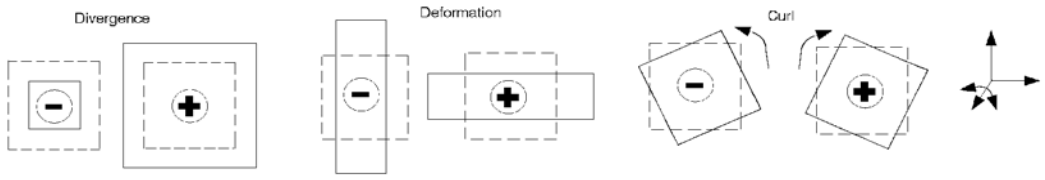


Figure 3.4 The figure illustrates the motion captured by the various parameters used to represent the motion of the regions. The solid lines indicate the deformed image region and the "-" and "+" indicate the sign of the quantity [Black and Yacoob 97].

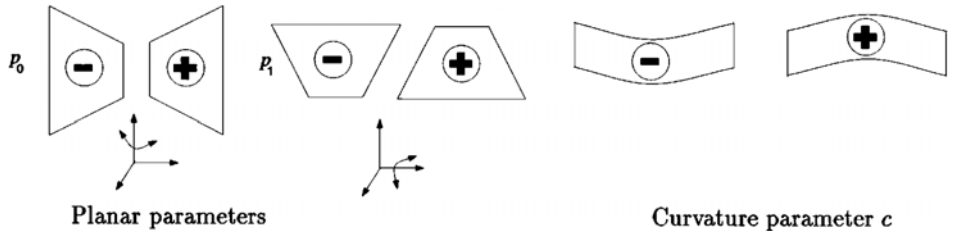


Figure 3.5 Additional parameters for planar motion and curvature [Black and Yacoob 97].

[Cohn et al. 98] track landmark points around facial features in image sequences by using Lucas-Kanade optical flow estimation method. By subtracting the position of features in current frame from the position of features in the first frame with a neutral face, the displacement of features were computed and used as predictors in discriminant function analysis. Separate discriminant function was used for different facial regions of brows, eyes and mouth. Finally separate group variance-covariance matrices were used for classification. Different from [Black and Yacoob 97], Cohn et al use Action Units in Facial Action Coding System (FACS) to represent facial expressions instead of using six basic emotions. 504 images sequences containing 872 action units or action unit combinations from 100 subjects are used in their experiments. Data were randomly divided into training and testing sets. A classification accuracy of 92 percent was reported for eyebrow region, 88 percent for eye region and 83 percent for mouth region. To make the system operates appropriately, the first frame must be expressionless and manually marked with landmarks. Also it suffers from the inaccuracy of optical flow based methods.

[Essa and Pentland 97] employ View-based and Modular Eigenspace methods [Pentland 94] to track the face in the scene and automatically extract the position of facial features such as eyes, nose and lips. Then optical flow is computed by using the method proposed in [Simoncelli 93]. This approach uses a coarse-to-fine continuous time Kalman Filter (CTFK) to obtain ‘correct’, ‘noisy-free’ motion estimation. Based on the flow estimation, they create a 2D spatial-temporal motion energy template to classify facial actions. In addition, they fit a general wireframe model to the input images to derive actuation of physical facial muscles in the paper (Figure 3.6). Also they point out the limitation of FACS representation of facial expression and propose an improved model - FACS+ which they believe could better capture the variety of natural human expression and model the temporal component of expression. They achieved recognition accuracy of 98 percent on 52 sequences of 8 subjects with either physical muscle model or 2D motion energy model.

[Matsuno 95] defines a Potential Net on a normalized face image. Potential Net is defined as a two dimensional mesh of which nodes are connected to their four neighbors with springs, while the most exterior nodes are fixed to the frame of the Net (see Figure 3.7). A node in the Potential Net is governed by a Partial Difference Equation which relates the position of the node to the gradient of a smoothed face image. They perform Principle Component Analysis (PCA) on the model nets obtained from 10 subjects with 4 types of facial expressions to construct the Emotion Space. Then test images are projected into this space and classified according to a simple nearest neighbor rule. The system achieves recognition accuracy of 80% on a small dataset with four expressions. The faces in the input image sequences must be in frontal view for this method to work. It's not reported in the paper how much training data is required to construct the complete Emotion Space.

[Wang etc. 98] utilizes a labeled graph of 19 Facial Feature Points (FFP) to represent faces (Figure 3.8). Twelve of them are used in facial expression recognition. The rest seven is used to preserve the local topology of the face. Each FFP is treated as a node of the labeled graph. The label of the node is defined as the template of 17*17 gray levels around the node. The links between nodes are empirically weighted with parameters related to facial characteristics. FFP tracking in the image sequences is obtained by solving a graph matching problem as proposed by [Buhmann et al. 1989]. As shown in Figure 3.9, the tracking system consists of two layers: a memory layer and an input layer. The memory layer contains the label graph from last frame and the input layer contains that of current frame. A correspondence is set up by minimizing a cost function using a simulated annealing algorithm. The cost function takes both the similarity and topological constraint between two labeled graphs into account. For each of three emotion categories – happiness, surprise and anger, they fit a B-Spline to the trajectory of each of 12 FEFPs. These B-spline curves constitute the standard templates for each basic emotion. The expression is classified according to the minimum distance between the actual trajectory of FEFPs and the trajectories defined by the template. The average recognition accuracy is 95 percent on a dataset consisting of 29 image sequences of three emotional expressions from 8 subjects. This method cannot deal with head motion and the first frame of input images need to be labeled by hand.

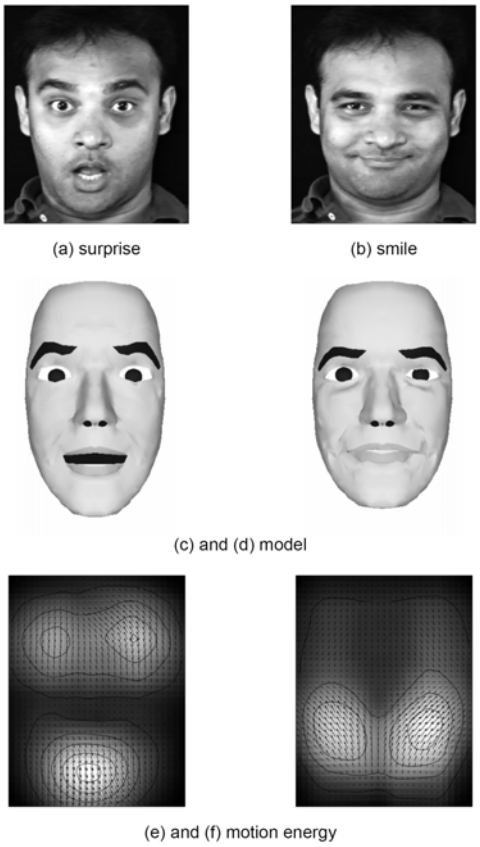


Figure 3.6 Determining of expressions from video sequences (a) and (b) show expressions of smile and surprise, (c) and (d) show a 3D model with surprise and smile expressions, and (e) and (f) show the spatial-temporal motion energy representation of facial motion for these expressions. [Essa and Pentland 97]

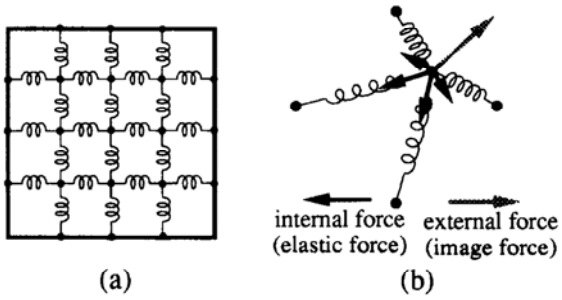


Figure 3.7 Structure of Potential Net and nodal deformation [Matsuno 95]

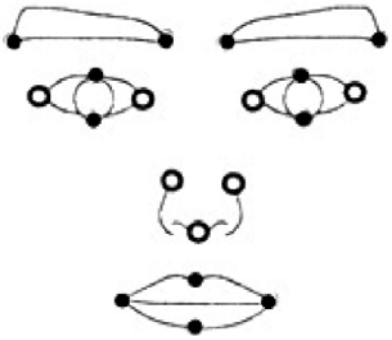


Figure 3.8 19 FFPs. [Wang etc. 98]

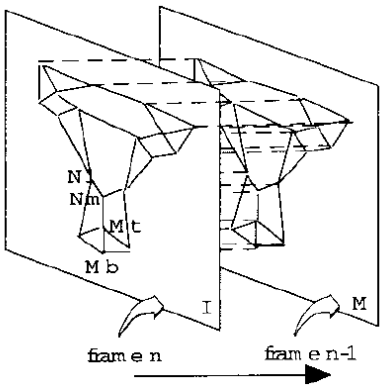


Figure 3.9 The FFP tracking system [Wang etc. 98]

Beginning from 2000, machine learning techniques (more specifically, graphical models) are widely used in facial expression analysis to better model the temporal behavior of facial expression and uncertainty caused by noisy measurements and ambiguity of facial actions.

[Lien et al. 97] used three different features extracted from input image sequences: dense optical flow, sparse feature points and lines and edges structure of the images to recognize AUs or AU combinations. Dense flow is computed through a coarse-to-fine Cai-Wang [Cai and Wang 96] wavelet motion representation. [Lucas-Kanade 81] flow estimation algorithm is used to track sparse features points across image sequences. And gradient filters are used to detect high-gradient components in the images such as transient wrinkles and furrows. These features are then quantified into a vector and fed into a discriminant analysis system or Hidden Markov Model (HMM) to recognize Action Units in the face images.

They tested their algorithm on Cohn-Kanade expression database and achieved an recognition accuracy of 92 percent by using dense flow extraction with HMM, 91 percent by facial-feature tracking with discriminant analysis, 85 percent by facial feature-tracking with HMM, and 88 percent by high-gradient component detection with HMM in brown region of the face. In the eye region, three action units (AU 5, AU 6, and AU 7) were classified with 88 percent accuracy with sparse feature points and discrminant analysis. For mouth region, 6 AUs are correctly recognized with the accuracy above 80 percent. Since each AU or AU combination is associated with a HMM, it is intractable to use it to recognize a large number of AU combinations (more than 7000 as reported in [Lien et al. 97]) in practice.

[Tian et al. 01] use three-state parametric model to represent different components of a face (see Figure 3.10). Transient wrinkles and furrows are detected using a Canny edge detector. After extracting parameters of these models during facial tracking, two separate three-layer neutral networks are used to recognize AUs in the upper and lower face. An AU combination in the upper or lower face is considered as a new AU in that part of the face. The system recognizes six upper face AUs with average recognition rates of 95.4 percent and ten lower face AUs with average recognition rates of 95.6 percent on Cohn-Kande and Ekman-Hager database.

Using four physiological signals (facial muscle tension, blood volume pressure, electrodermal activity of skin and Hall effect respiration around the diaphragm), [Picard et al 2001] derive six statistical features and ten physically-motivated features. They use a Hybrid Sequential Floating Forward Search with Fisher Projection (SFFS-FP) algorithm to select features and classify the signals into one of eight categories: *no emotion, anger, hate, grief, platonic love, romantic love, joy, and reverence*. They captured physiological data from a particular subject for 30 days and investigated the problem of day-dependence of human emotion. They found out from the experiments that physiological features for different emotions from the same day tend to be more similar than features for the same emotions on different days.

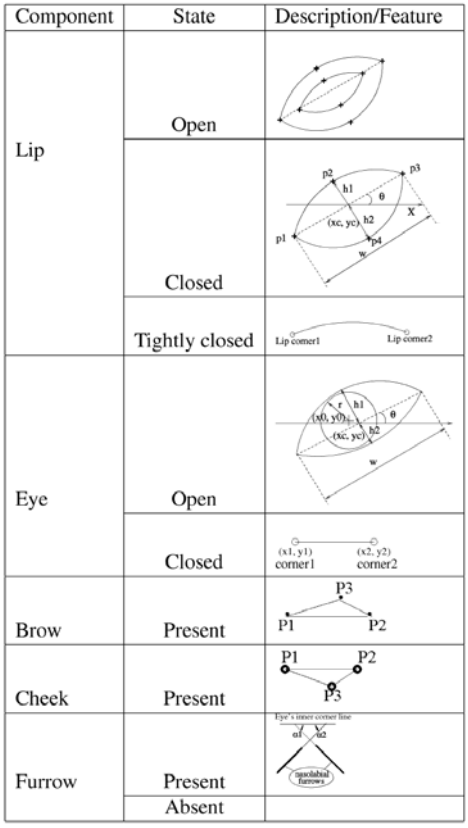


Figure 3.10 Multi-state Facial Component Models of a Frontal Face [Tian et al. 01]

[Cohen et al. 03] investigate the expression recognition problem by using a Naïve-Bayes classifier with Cauchy distribution. A Tree-Augmented-Naïve Bayes (TAN) classifier is used to learn the dependencies between facial features. To better describe the dynamic nature of expressions, they use a multi-layer HMM model (Figure 3.11) to automatically segment and recognize long expression video. Each sequence of the video was first fed into a lower level HMM to produce a sequence of state variable from each of six separate HMMs trained for each basic emotion. These sequences were then fed into a higher level HMM which models the transition between different expressions to segment the video and recognize the expression simultaneously. An approximately 90 percent accuracy is achieved for Happy and Surprise expression on Cohn-Kanade database. This work is the first to study the dependencies between facial features in expression recognition and is also the first to segment a long video sequence and recognize facial expression simultaneously. However, expression recognition has to begin from a neutral face. The transition from one expression to another must pass through the neutral state too. This assumption may not hold in practice.

[Gu and Ji 04] presented a system to monitor driver vigilance. They utilize cues from multiple sensors and robustly track eye pupils, facial landmarks, furrows and head motion. Facial AUs are identified from these tracking results base on a rule based description. Then a Dynamic Bayesian Network (DBN) was used to infer the driver's vigilance level (Figure 3.12). In addition, Another DBN was defined in terms of *phase instance.* One phase instance is a continuous facial change which proceeds for a limited period of time covering a single ONSET, APEX or OFFSET duration. This phase based DBN is used to identify not only the vigilance level (Inattention, Yawning, Falling asleep) the driver is currently in but also the phase (Detection or Verification) of the vigilance level. The temporal evolvement of facial expression is described by a first-order Markov model in the DBN. For each input frame, the bottom-up inference is used to classify current facial display. Base on current classification result, the top-down inference is conducted on phase based DBN to actively select effective visual channels from pre-learned curve of posterior probability. This makes the system robust to the ambiguity of facial expression and failure of one sensing channel.

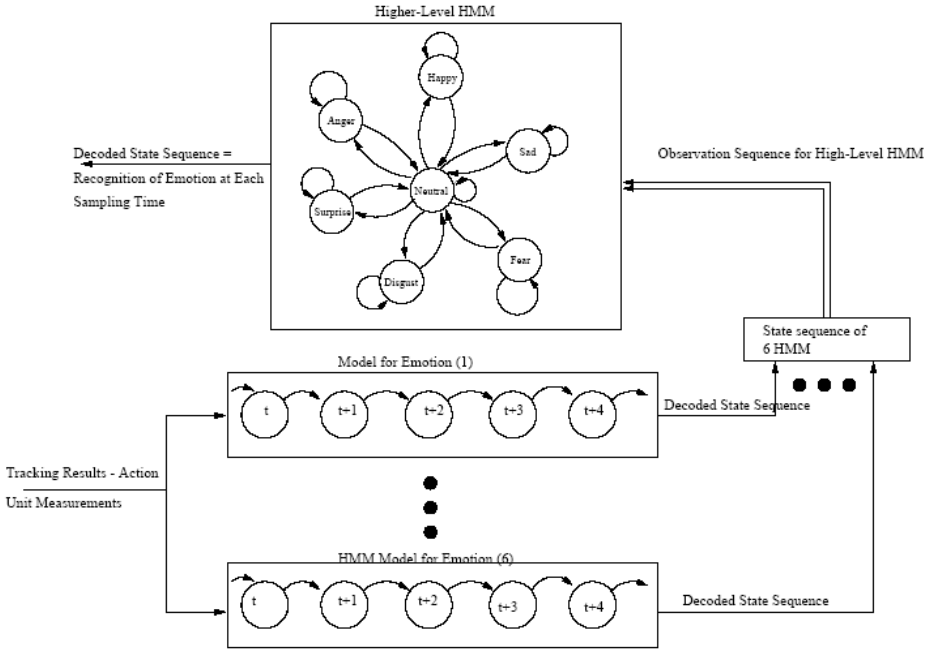


Figure 3.11 Multilevel HMM architecture for automatic segmentation and recognition of emotion [Cohen et al. 03]

[Zhang and Ji 05] use the similar idea to recognize facial expression based on detected AUs. They use an IR eye tracker to determine the position of pupils first. Then 26 facial features around the eyes, nose and mouth are tracked using Kalman filtering

based on the position of eyes detected previously. Similar to [Gu and Ji 04], some predefined heuristics are used to infer the AUs from detected positions of facial features. These AUs are then fed into DBN to perform expression classification. This system could deal with subtle head motion (<30 degree) and is robust to certain degree of partial occlusion. Figure 3.13 shows the recognition result from an occluded face. We could see even if the face is partially occluded by hands of the subject, the system still could correctly classify the expression.

Motivated by the observation that facial action detection are highly related to facial expression recognition, [Dornaika and Davoine 05] propose to use a particle filter to simultaneously track facial actions and recognize expressions. They first recover the 3D head pose using a deterministic registration technique based on Online Appearance Model. Then the coefficients of deformable face model are concatenated with a discrete variable which is the label for one of six basic emotion categories, to construct a state vector. This vector is then recovered in the tracking algorithm using a particle filter. For each expression, they train a second order Markov model to describe the dynamics of the expression.

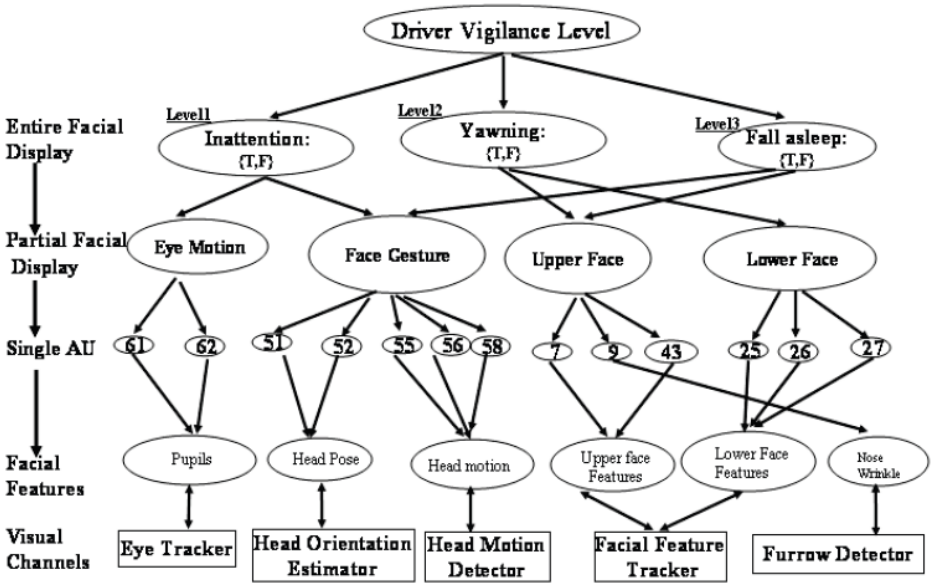


Figure 3.12 BN for vigilance detection [Gu and Ji 04]

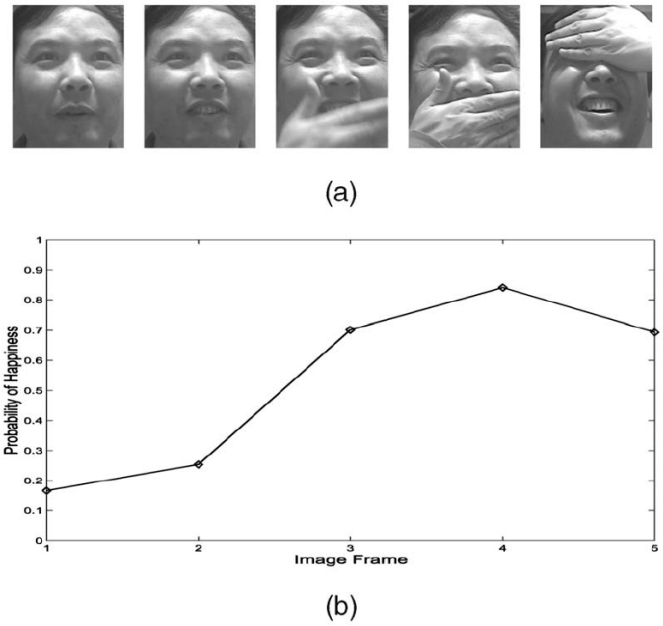


Figure 3.13 (a) A posed image sequence performing occluded facial expression. (b) The result from our facial expression model. [Gu and Ji 04]

Tensor (multi-linear) model was first introduced to the field of face recognition and analysis by Vasilescu and Terzopoulos in 2002 [Vasilescu and Terzopoulos 02a, 02b], then used in a broad range of areas such as super resolution of face images, texture synthesis [Vasilescu and Terzopoulos 2004] and facial animation [Vlasic etc. 05]. It is a generative model which elegantly deals with multiple varying factors in face modeling or analysis. A tensor is high-dimensional generalization of the concept of a matrix. Higher order singular value decomposition (HOSVD) and tensor model could be used as an effective dimension reduction technique which has been proven to outperform Principle Component Analysis (PCA) model in face recognition[Vasilescu and Terzopoulos 02a]. But it is not clear if facial dynamics could be modeled using a tensor framework before Gralewski et al.'s work [Gralewski et al. 06] was presented. They use a 4^{th} tensor to represent the entire data. The four dimensions of the tensor are number of sequences, the dimensionality of the point space, and the number of control points and the length of the input sequence. By performing HOSVD on the tensor model created from image sequences, they conclude that the motion signatures extracted from HOSVD encode information about gender, emotion and identity. However, no quantitative analysis and experiments are proposed in the literature. And they didn't test the algorithm on a publicly available database either. Also the control points were manually labeled from video footage to set up the correspondence between frames. This is time-consuming and

infeasible for long video analysis in practice. An automatic tracking algorithm is essential to the improvement of this work.

On the other hand, non-linear dimensional reduction techniques such as ISOMAP [Tenenbaum 00] and local linear embedding (LLE) [Roweis 00] are promising in dealing with high-dimensional data. Manifold based analysis of facial expression was presented by Chang and Turk et al. [Chang 2005]. Each face image is represented by 58 control points corresponding to an Active Shape Model (ASM) [Cootes 95]. The original video is embedded into a 3-dimensional manifold using a modified Lipschitz embedding algorithm. After the embedding, a complete expression sequence becomes a path stemming from a center which corresponds to the neutral expression on the expression manifold (Figure 3.14). For each cluster of expression in the embedding space, an ASM model is trained. Then they use ICondensation to track facial features represented by ASM and perform expression classification simultaneously in a common probabilistic framework. The system assumes constant illumination and near front view face pose. And it doesn't model facial details like wrinkle and furrows. More importantly, it's not clear why they chose to map original video into a 3-dim manifold. i.e. Is human expression space is indeed 3 dimensional? These problems are still need to be solved to obtain better results.

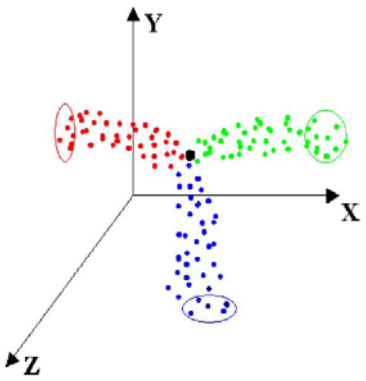


Figure 3.14 Illustration of a 3D expression manifold. The reference center is defined by the neutral face. Image sequences from three different expressions are shown. The further a point is away from the reference point, the higher is the intensity of that expression. [Chang 2005]

Similar to [Zhang and Ji 05] and [Gu and Ji 04], [Tong and Ji 06] use dynamic Bayesian Network (DBN) to describe the relationship between different AUs, while in previous work all the AUs are assumed to be independent to each other. Figure 3.15 demonstrates two BNs consisting of 14 AUs. The one shown in (a) is a prior BN derived

by a human expert and (b) shows the one learned from training samples. We can see the learnt BN keeps all the relationships identified in (a) but adds some links to it (e.g. the link from AU27 to AU6). These links represent the relationships in the training data but neglected by human experts. An average recognition rate of 93.33 percent is achieved with this algorithm on Cohn-Kanade database. Since the method presented here uses a holistic Gabor representation, it will require robust and precise image alignment.

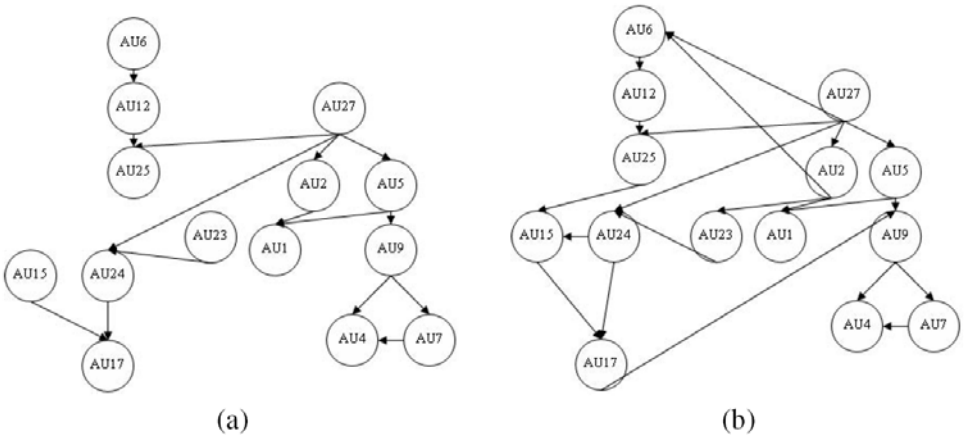


Figure 3.15 (a) The prior BN for AU modeling before learning; (b) The learnt BN from the training data [Tong and Ji 06]

[Kanaujia and Metaxas 06] use Non-negative Matrix Factorization (NMF) instead of PCA to train an ASM which could efficiently characterize localized deformation of facial features. By using localized NMF (LNMF), the model preserves detailed local deformation accurately. Conditional Random Field (CRF) is used to learn the dependencies between facial features and recognize expressions directly. More than 93 percent correct predictions are reported for the expressions of Surprise, Sadness, Fear and Joy on Cohn-Kanade database.

Cohn et al. [Cohn and Kanade etc. 01] (CMU/Pittsburgh group) solve the AU recognition problem for spontaneous expression in a different way. They first tracked faces and recovery head motion from image sequences. Based on the head motion, the face images are stabilized. Then blinks and AUs in brown region of the face are detected according to a heuristics based description of AUs. Frank-Ekman database is used to test the system. An overall 98 percent accuracy was reported for blink detection, while only 57 percent was achieved for AU detection in brown region. The authors claimed the lower accuracy in the brow region was due to low intensity of AUs in the brow motions.

[Bartlett et al. 06] (UCSD group) present a fully automatic AU recognition system

for spontaneous expression analysis. The system uses real-time face detection system presented by [Viola and Jones 04] to locate faces automatically. Then it constructs a feature pool by filtering normalized images using a bank of Gabor filters with 8 orientations and 9 spatial frequencies. The features are selected from this pool through AdaBoost method for each of 20 AUs. Then the input images are classified using SVM based on these features selected by AdaBoost. A mean recognition rate of 91 percent is obtained on Cohn-Kanade database and 93 percent on RU-FACS-I database. The output margin of the classifier describes the intensity of AUs.

3.4 Analysis-by-Synthesis

The broad and in-depth application of Computer Graphics in Computer Vision field is a trend in the past few years. A good example is analysis-by-synthesis technique. Facial expression modeling and animation has been an active research area in Computer Graphics for a long time. Staring from Parke's pioneering work [Parke 72], graphics researchers try to find a simple but efficient tool to easily create facial animation. In early days, lots of work [Waters 87, Waters and Terzopoulos 91, Lee and Terzopoulos 95, etc.] made great effort to build a physics based simulation of facial muscles. While it turns out to be difficult to simulate the facial muscle system of humans with great fidelity at relatively low costs, there are still research groups [Sifakis etc. 05, 06] working on simulating complex facial muscles and creating high-quality facial and speech animation based on finite element and level set techniques. On the other hand, data-driven approaches [William 90, Pighin et al. 98, Blanz and Vetter 99, Noh and Neumann 01, Wang and Samaras 04, Vlasic and Popovic 05, etc] appear to be more promising and easy-to-use than physically based methods. Data-driven approaches utilize motion capture data to track facial deformation and create a parametric representation of facial motion. Then this parametric model could be either used in analysis or synthesis of new expressions. See [Pighin and Lewis 06] for a comprehensive survey of facial expression modeling and animation techniques.

3.5 Evaluation of Expression Analysis Algorithms

While a large amount of expression analysis is presented in the past decades, there is no systematic study about the evaluation of these algorithms Different recognition accuracy was reported on different dataset. Also the testing procedures are different either. This makes rigorous comparison between different algorithms impossible. Obviously, the lack of a standard evaluation system will hinder the development of future expression analysis system. A publicly available database associated with standard algorithm testing

procedures similar to FERET [Phillips etc. 00] would be helpful to improve current expression analysis system. Cohn-Kanade database seems to be a good candidate for this purpose. However, the deliberate expressions elicited in controlled environment contained in this database don't capture the richness and ambiguity of spontaneous expressions. Factors complicating the expression analysis in practice, like head motion, occlusion and speech related expression, are not well captured in the database either. These simplicities make Cohn-Kanade database not suitable for the evaluation purpose. From this point of view, RU-FACS-I could be the first step toward our objective. See Section 2 for a detailed description of these two databases. Given a public database, how to divide the expression data into different training and testing sets is another important question. The division should be able to test the generality of the algorithms as well as their robustness to different level of complexities of expression analysis. This remains another direction of our future work.

4 Analysis and Visualization of 3D High Resolution Expression Data

In this section, we try to answer following questions:

1. Is it possible to track subtle expressions and calculate dense motion flow from them using 2D optical flow algorithms?
2. Is the motion flow calculated in step 1 correct? Or how to verify the flow calculation results using dynamic high resolution 3D data?
3. If we do get correct flow, how to visualize and analyze them?
4. When does the 2D optical flow algorithm fail? And how can we improve it using our 3D data?

4.1 Data Preprocessing

Dynamic high resolution 3d expression data was captured at 30 fps with a structured-light 3d scanner. After de-noising, hole-filing and other preprocessing steps, we obtain clean dynamic facial expression data for our experiments. Figure 4.1 shows the screenshot of one sample frame from the sequence.

Figure 4.1 One sample frame

4.2 Image Data for Flow Computation

Ideally, we could use the texture images extracted from our 3D data for optical flow computation. Unfortunately, the available 3D data contains only per vertex color value (not texture coordinates). So we should not use original texture images if we hope to know the correspondence between the 3D data and the 2D image in order to verify the 2D flow result we get. As a result, we render the 3D data to the frame buffer and save the rendering results. These results are then inputted to the optical flow algorithm. Figure 4.2 and 4.3 shows two neighboring frames in a smirk expression sequence. You could find that there is little difference between these two frames. From the dense flow we get, however, you would see the clear motion pattern between these two frames.

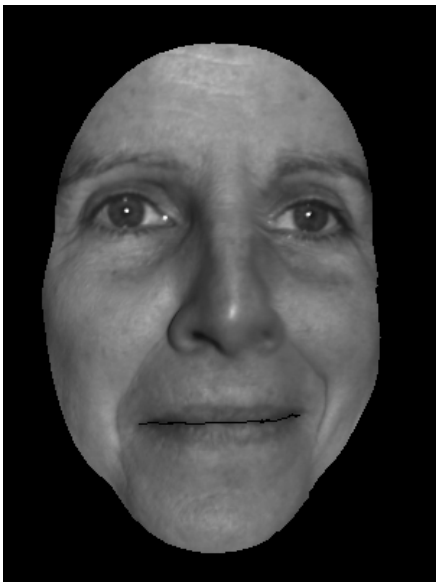


Figure 4.2 Frame No. 294

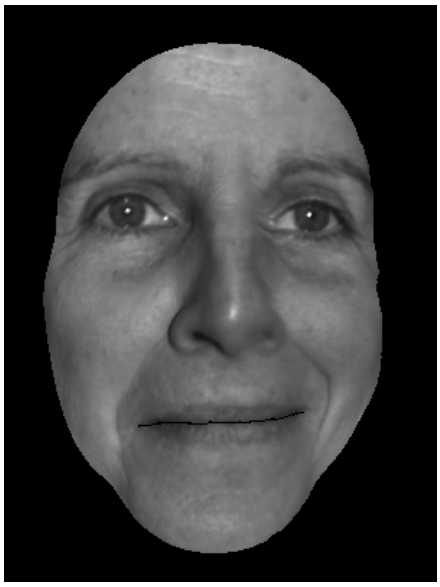


Figure 4.3 Frame No. 298

4.3 Optical Flow Computation and Visualization

We use the well-known dense optical flow algorithm [Black and Anandan 96] in our flow computation for its robustness and accuracy. A four level image pyramid is used and we accept all the default parameters in the algorithm. Once we calculate the dense flow, we use the Linear Integral Convolution [Cabral and Leedon 1993] method to visualize it. The result is shown in Figure 4.4. Different colors are mapped to different magnitudes of the flow with gray corresponding to zero and pure green [0, 255, 0] corresponding to

maximum flow magnitude. From the result, we could find out clear motion pattern around the left corner of the mouth. As we expect, the left mouth corner of the subject is pulled up and left as she is smiling.

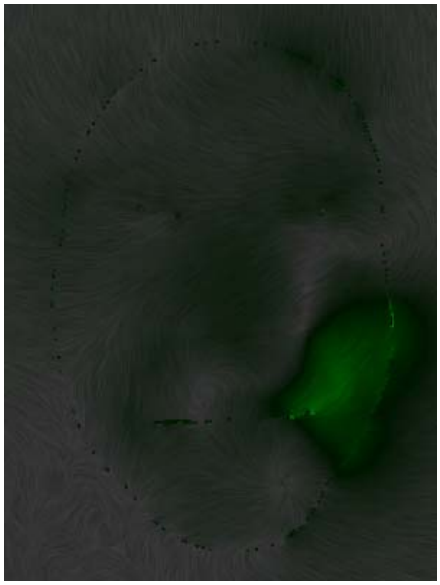
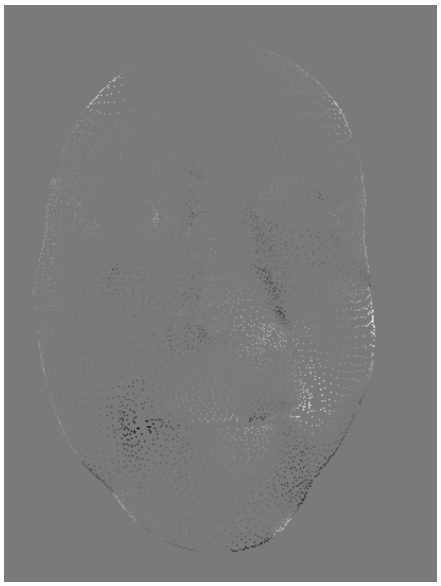


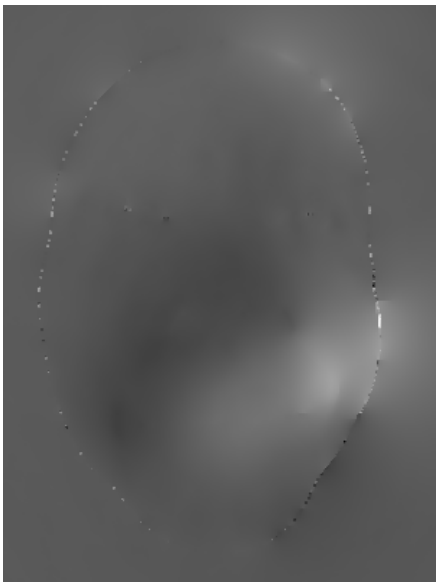
Figure 4.4 Flow result: 294->298

4.4 Flow Result Verification

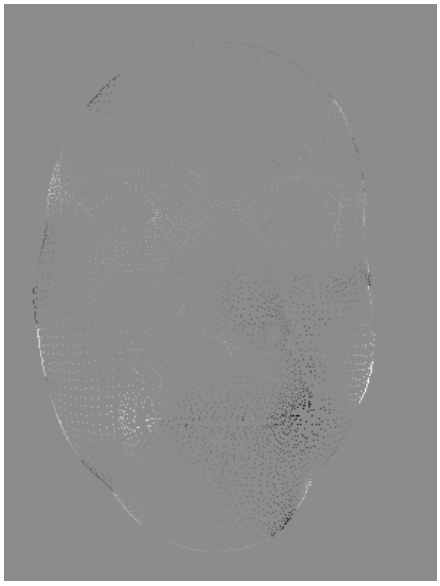
Before we make any conclusion based on the dense flow result, we need to verify the flow we get in the above experiment. The ground truth data we use to verify our dense flow comes from the 3D tracking result of [Wang and Gupta et al 05]. We project 3D vertices of the tracked face onto the same imaging plane as we used when rendering 2D texture images. Then we extract the ground truth flow by differentiating two projections of the same 3D vertex. The comparison between the ground truth flow and our flow result is demonstrated in Figure 4.5. (a) shows the u component of the ground truth flow and (b) is the v component of the same flow. (a') and (b') shows the corresponding component in our flow result. White dots in (a) or (b) represent strong flows along positive u or v axis. Similarly, black dots are strong positive flows. Areas not covered by ground truth data are filled with gray background. Due to rounding error in ground truth data generation, two flows are not quantitatively identical. But comparing two flows, you could find that they are in accordance with each other. For instance, in both flows there are strong negative v component and positive u component around the left mouth corner, which is in agreement with our observation of a smirk expression.



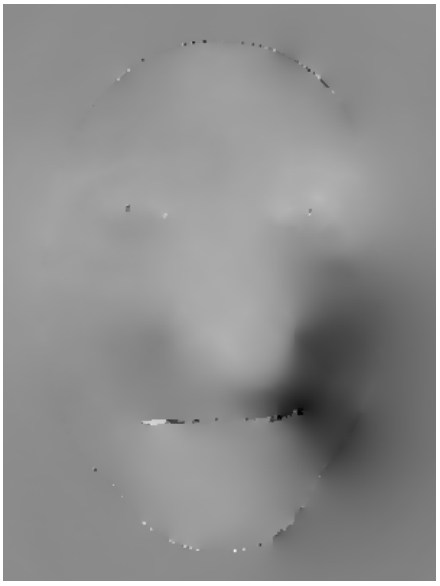
(a) ground truth: u



(a') our flow: u



(b) ground truth: v



(b') our flow: v

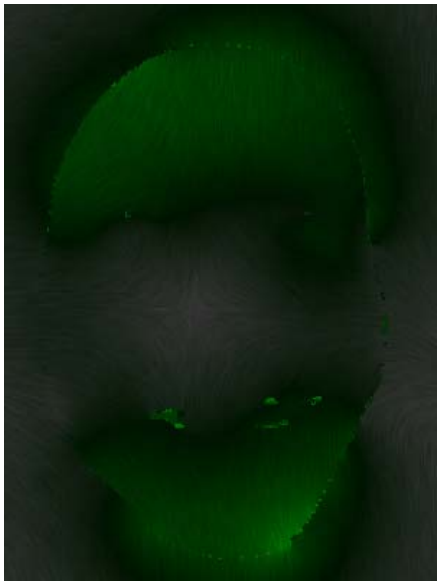
Figure 4.5 Comparison between ground truth data and optical flow result

4.5 Failed Cases and Why

Figure 4.6 shows one case when our algorithm failed to compute correct flow. In this case in and out of plane head motion seems confuse our flow computation algorithm a lot.

We are investigating the possibility of stabilizing the images using the rigid transformation computed from dynamic 3D data.

(a) Frame No. 370 **(b) Frame No. 374**



(c) Flow result: 370 -> 374

Figure 4.6 A failed case when head motion is introduced in the expression data

5 Expression Analysis by Using High Speed Video Camera

From previous analysis, we know high resolution 3D data could benefit the analysis of expression subtlety. But the algorithm fails when the data contain out-of-plane head motion. In this section, we try to overcome this problem by using data from a high speed video camera. When facial expression data are captured at 1000 frames per second or higher speed, head motion between two successive frames is not significant thus gives better results for the dense flow computation. More importantly, as these data capture more subtle changes with greater temporal resolution, we could analyze facial expression in a smaller time scale and discover different motion patterns from different people. Our algorithm could extract dense expression flow from the data correctly. And the visualization and analysis of the flow shows clear temporal patterns during different phase of the facial expression.

5.1 Data Set

The expression data we use in this section was captured by a high-speed Phantom v9 video camera. An experiment participant was instructed to make a “surprise” expression which was captured at 1000 fps with a resolution of 768 by 768. The whole onset, apex and offset phases together with blinks were recorded during video capture. Figure 5.1 shows a sample image from the video sequence.

Figure 5.1 One sample frame from a surprise expression

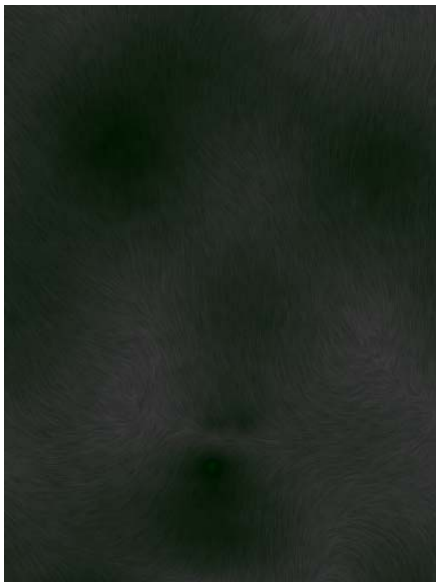
5.2 Flow Computation

Before we compute the dense flow from the video data, we down-sample the video into 30 fps for two reasons. First of all, we try to verify if our algorithm could extract correct flow information from a regular 30fps video sequence. Second, the deformation between two neighboring frames in the original video data may be not significant enough to be detected by the flow computation algorithm. That's to say, the computed flow will be too noisy if we use the original video directly. Similar to the previous experiment in Section 4.3, we use the dense optical flow computation algorithm [Black and Anandan 96] in our flow computation for its robustness and accuracy.

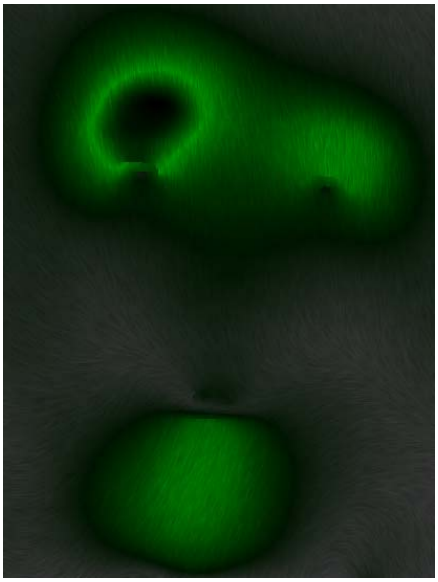
5.3 Flow Visualization

After we calculated flow data, we visualize the flow using Line Integration Convolution (LIC) as we did in previous experiments. Parts of the results are demonstrated in Figure 5.2.We could easily identify different phases of the expression by simply looking at the visualization results. In Figure 5.2 (a), the subject doesn't move much from frame 118 to 119, which means that the subject is in the same phrase of an expression. In this case, the subject is in the apex phrase of a surprise expression. Figure

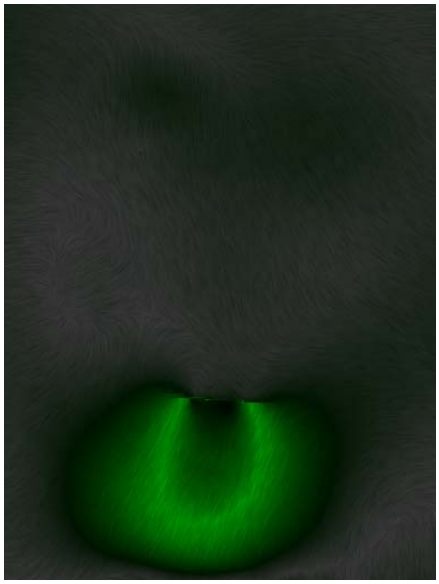
5.2 (b) shows significant deformation around both eyes and mouth, which is a strong signal of phrase transition from the apex to the offset of the surprise expression. Similarly, in Figure 5.2 (c), large vertical movement around the mouth area shows that the subject keeps his eyes open but is trying to close his mouth. Last but not least, we could easily find out the subject is blinking from the flow demonstrated in Figure 5.2 (d).



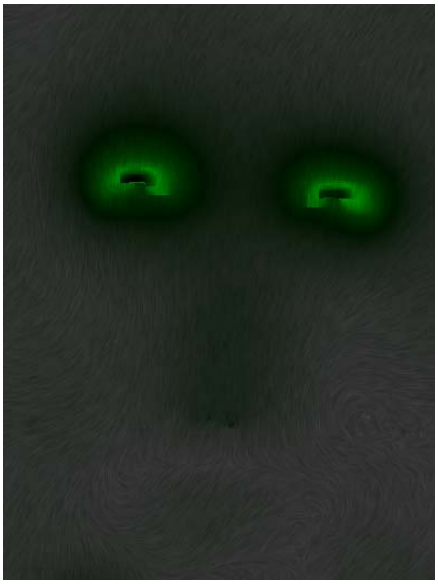
(a) Flow result: 118 -> 119



(b) Flow result: 122 -> 123



(c) Flow result: 126 -> 127



(d) Flow result: 134 -> 135

Figure 5.2 LIC visualization of the flow results

Figure 5.2 shows that LIC is a very good way to visualize our dense flow results. However, it can't clearly visualize the dynamic nature of the flow. i.e it is not intended for time varying flow visualization. For the sake of describing temporal evolvement of the expression, we need another way to express dynamic aspect of our flow data. Image Based Flow Visualization (IBFV) [Wijk 2002] creates an animation of a sequence of images which are warped based on the flow data and blended with background images from frame buffer. It's well-known for its simplicity, efficiency and comprehensive support for a wide variety of visualization techniques. We use IBFV in our experiment to create dynamic visualization animation for our flow results. One screen shot of the result is shown in Figure 5.3. The complete animation video is available in the same folder as this thesis submission.

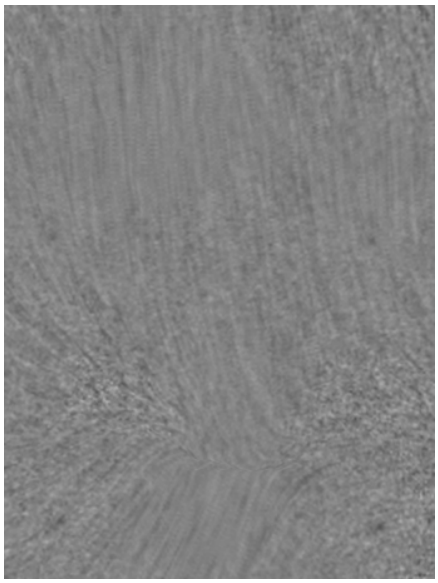


Figure 5.3 Snapshot of IBFV results

5.4 Statistical Analysis of Flow Data

In section 5.3, we demonstrate how flow visualization techniques help us discover motion patterns in an expression sequence. In this section, we try to perform some simple statistical analysis on the flow data and show results of quantitative temporal analysis. We define three regions around two eyes and the mouth. Foe each of these three regions, we find the max deformation within the region. Plotting this for all frames in the video, we get Figure 5.4 in which the x axis is the frame number and y axis shows the maximum deformation within each region in terms of pixels. Investigation of Figure 5.4 reveals that

from frame No. 22 to frame No. 25, both the subject's mouth and eyes are moving significantly. From frame No. 26 to frame No. 30, however, the mouth (jaw) keep moving while there's not much deformation around eyes any more. One of the reasons for this is that the mouth/jaw has more degree of freedom than eyes and could have larger deformation than eyes. More interestingly, we could see a clear burst of eyes' deformation while that of the mouth keep unchanged from frame No. 33 to frame No. 38. Obviously, this signals a blink during the timeframe. If we look at the result from frame No. 22 to frame No. 25 more carefully, we could figure out that the right eye of the subject moves faster than the left eye. These experiments demonstrate that our algorithm could be used to qualitatively and quantitatively describe the motion patterns hidden in the expression data.

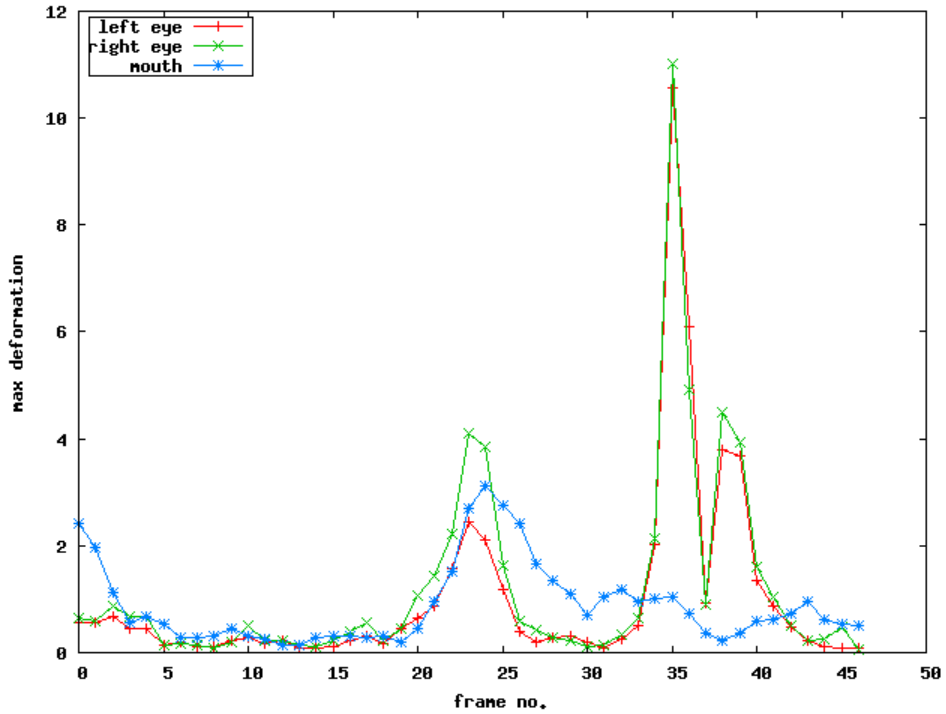


Figure 5.4 Deformation statistics around eyes and mouth

5.5 Comparison of Flow Results from Different People

In this section, we try to visualize and compare the flow results from different people. We captured six universal expressions from five different subjects. Two of them are male and the rest three are female. Ethics of the subjects are European and Asian. All

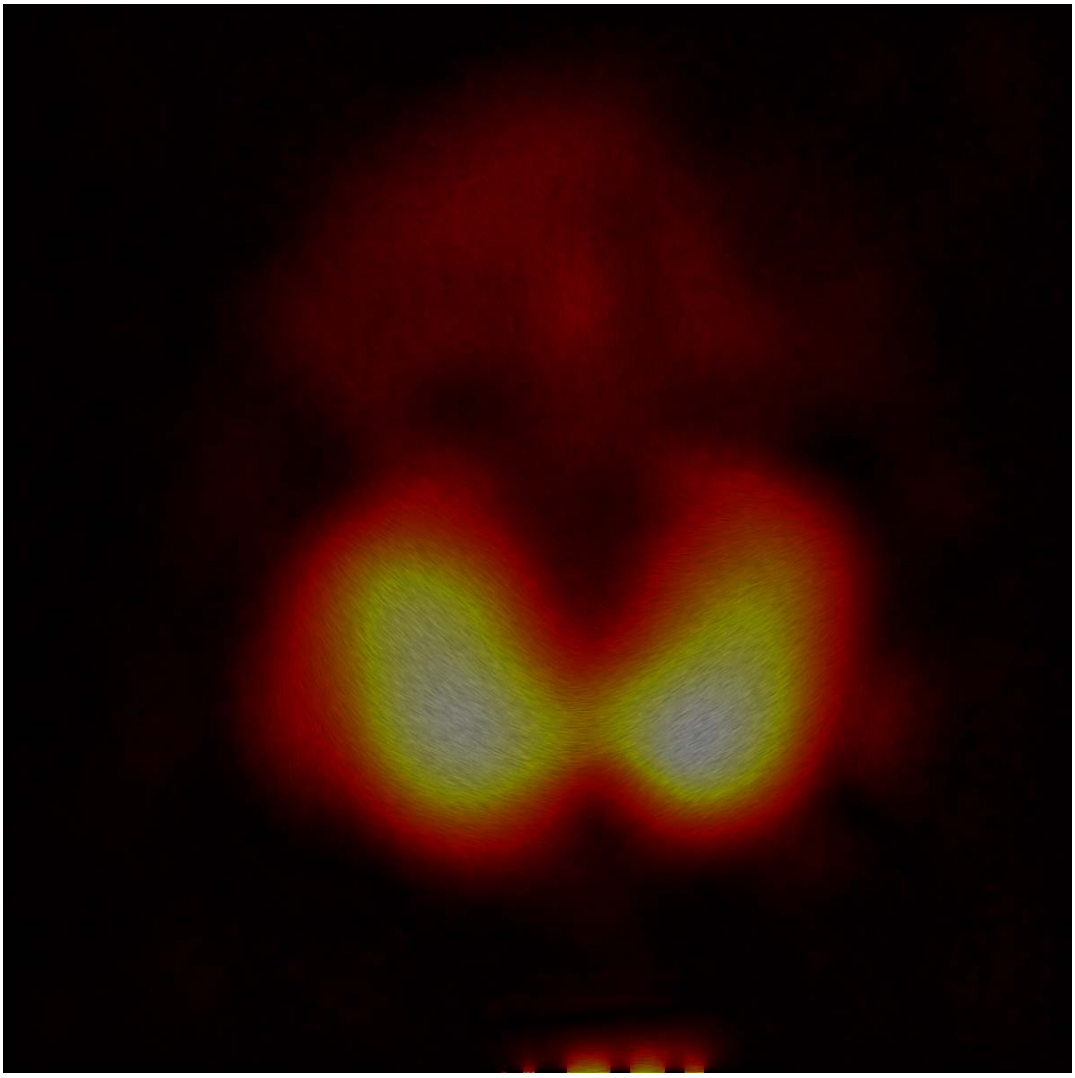
expressions are captured by a Phantom V9 video camera at 300 frames per second with an image resolution of 1200 by 1200. During the experiment, subjects are instructed to make six different universal expressions -- happiness, surprise, anger, sadness, fear, disgust.

Dense flow fields are computed from the expression video sequences and visualized using our tool. Figure 5.5 shows one sample frame from each flow result computed from a smile video sequence. In this figure, flow vectors are mapped to different colors based on their magnitude. Black represents the minimum value of the vector magnitude and white the maximum. Starting from black, the color mapping goes smoothly through red, orange and yellow to white.

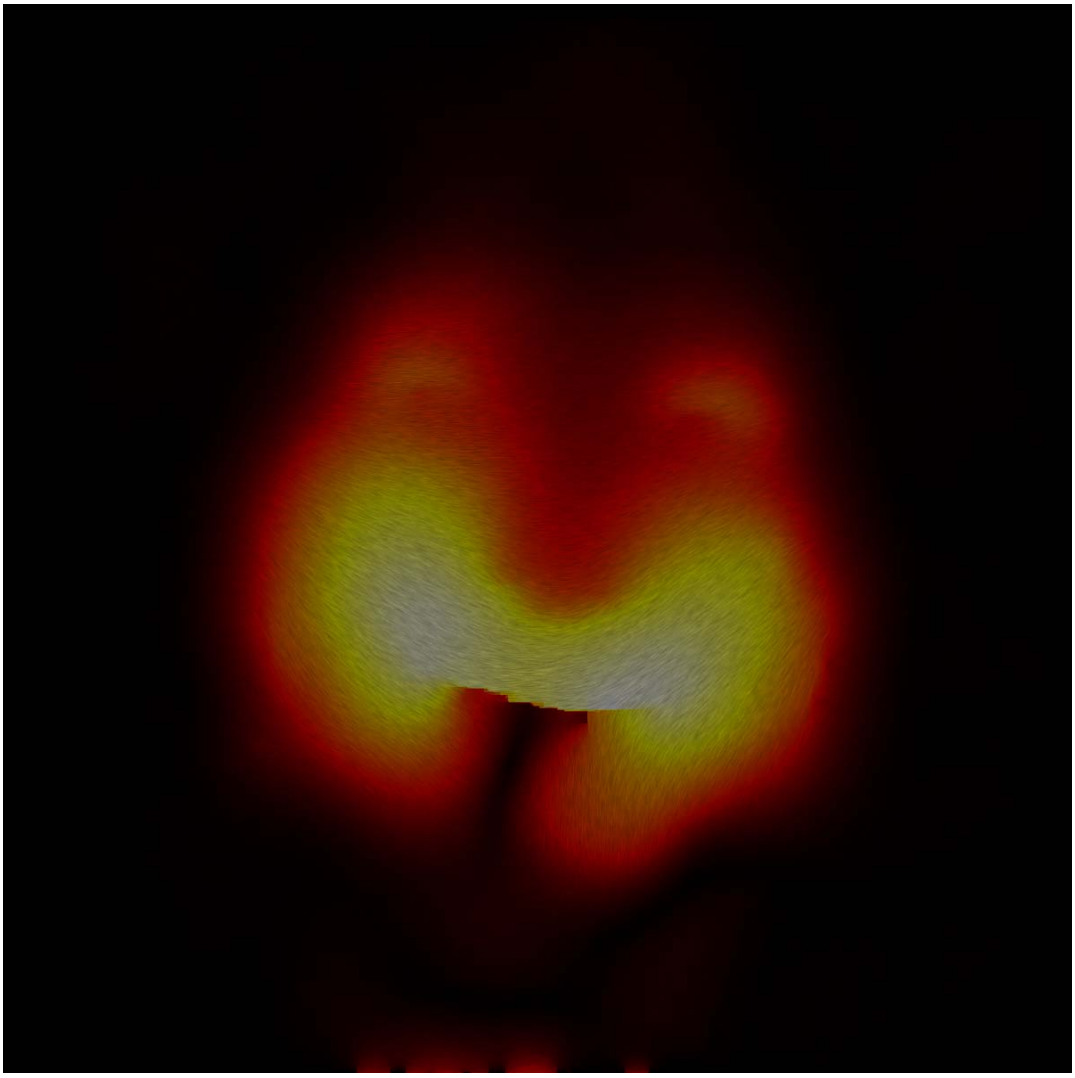
We could tell, from Figure 5.5, that there exists a common motion pattern for all these flow results as they are computed from the same expression. Centered at the corner of the mouth (the white part of each image), two circular structures correspond to the motion of two cheeks of the face. In the rest of this thesis, we call these circular structures Significant Deformation Area (SDA). As we expected, two corners of the mouth demonstrate the most significant motion in a smile expression which is controlled by the contraction of Zygomatic Major. Away from the mouth corner, the deformation decays gradually until it reaches the boundary of an SDA.

While all these flow results demonstrate similar patterns for the same expression, each individual shows its personality when making the same expression. First of all, the SDAs showed in the image have different area and shape. Subject No. 3 and 4 have a larger SDA than those of others, which is in accordance with the fact that these two subjects are males and make more exaggerated expressions. The shapes of the SDAs are different as well. e.g. Subject 1 shows a nearly circular SDA while Subject 3 has an SDA of ellipse shape with large eccentricity. Secondly, the direction of the motion vectors within an SDA is different. Subject 3 and 4 show dominated vertical motion while the rest of subjects have more horizontal motion. This is due to the difference on geometry of the faces and their underneath muscle configurations. Last but not least, the SDAs demonstrate symmetry and discontinuity of the expression. Each subject demonstrates different asymmetry of the same expression and some subject shows motion discontinuity around the mouth area.

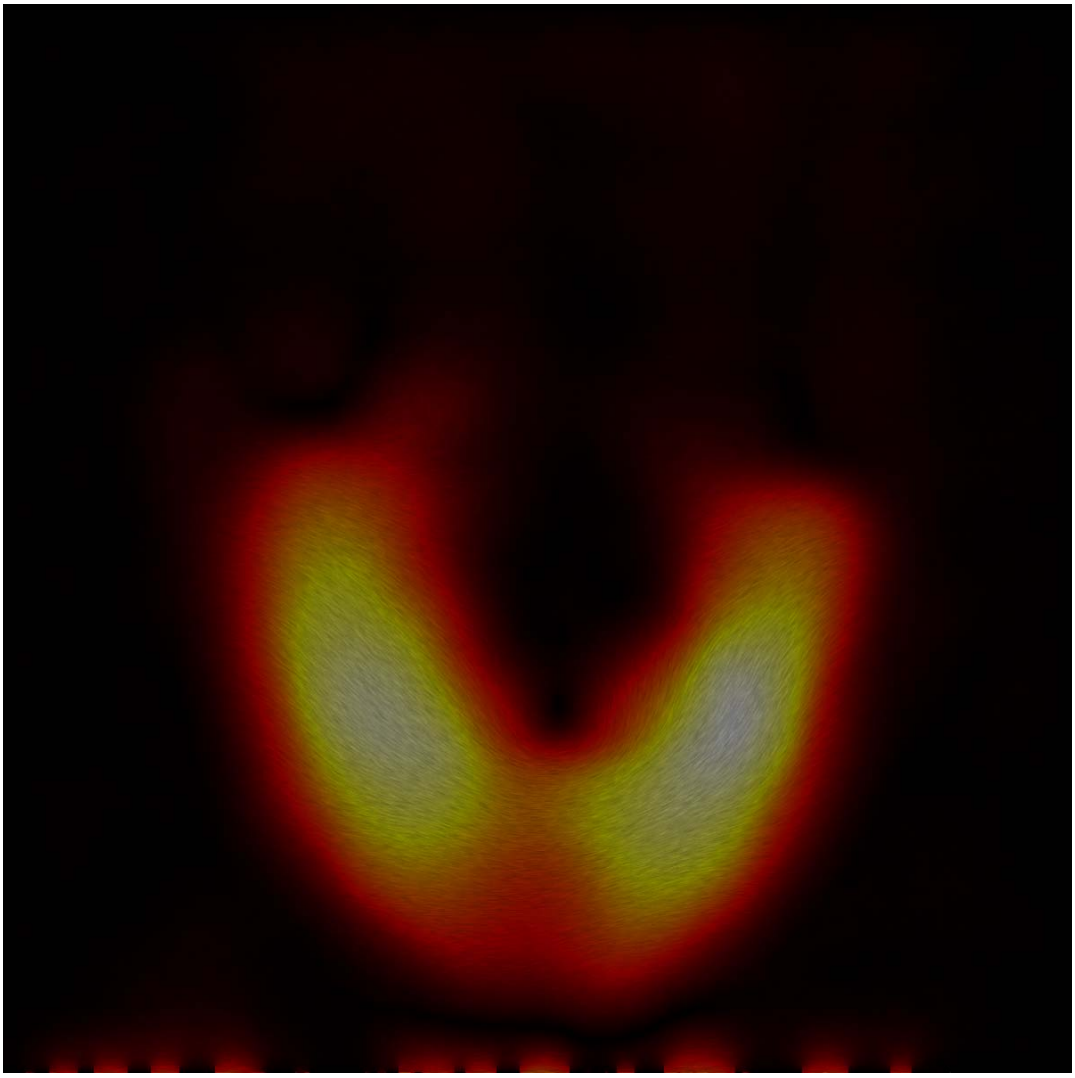
In summary, we develop a tool to compute dense flow from dynamic expression data, visualize the flow filed using LIC and color mapping and analyze the SDAs of different flow results. Experimental results demonstrate common motion pattern for the same expression across different subjects while each individual shows different characteristics within the same motion pattern.



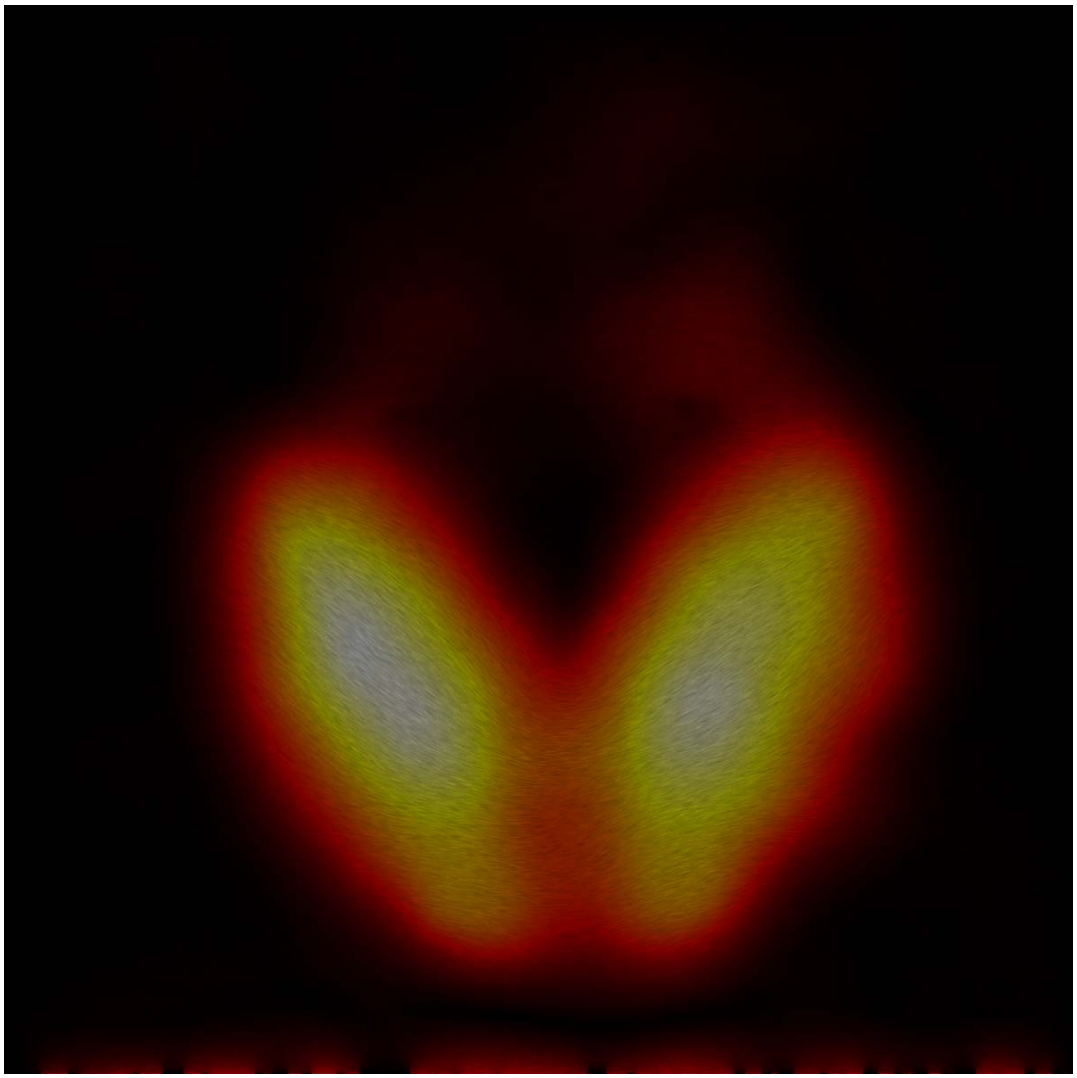
(a) Flow visualization of subject No. 1



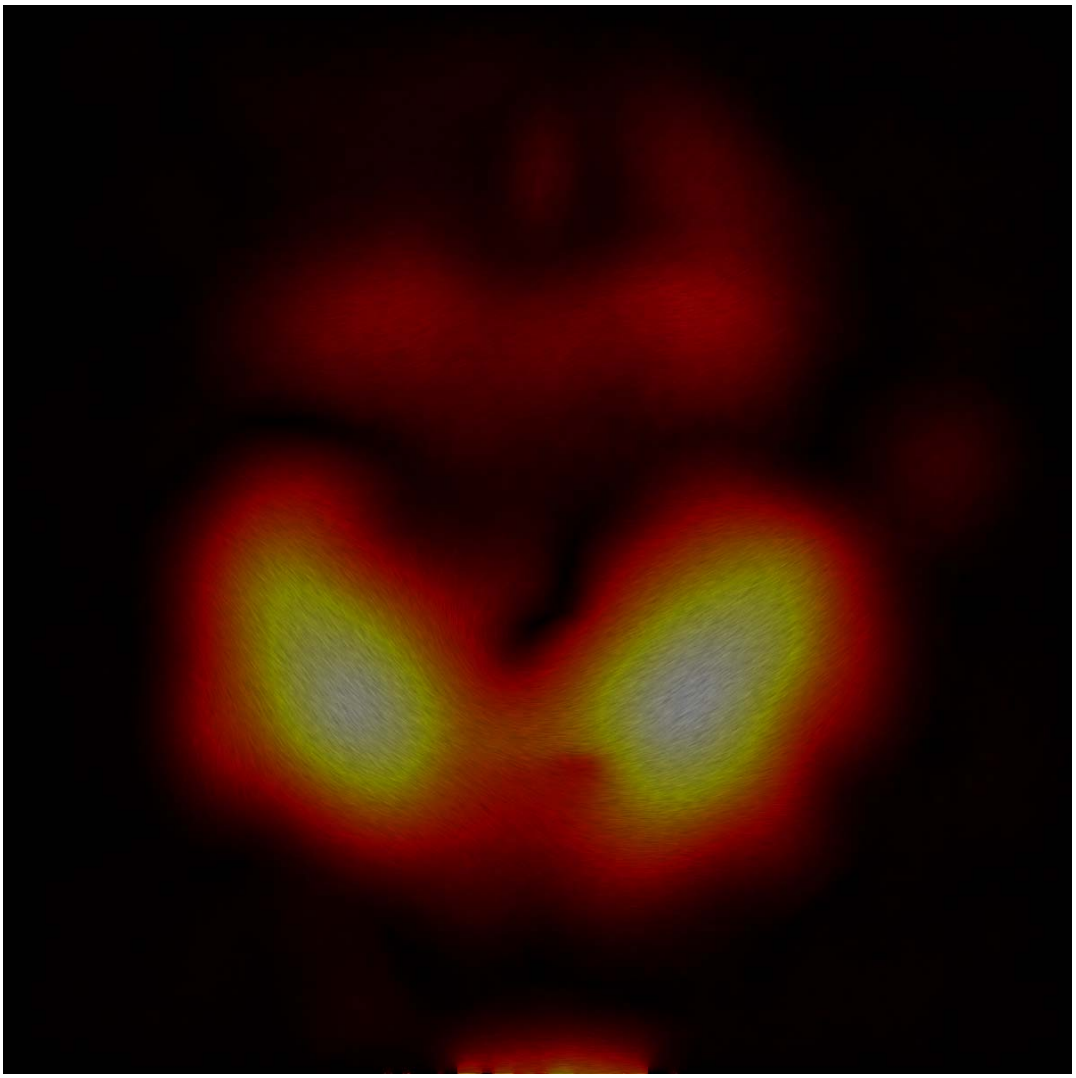
(b) Flow visualization of subject No. 2



(c) Flow visualization of subject No. 3



(d) Flow visualization of subject No. 4



(e) Flow visualization of subject No. 5

Figure 5.5 Comparison of different flow results

6 Summary

In this paper, we reviewed the techniques for automatic spatial-temporal analysis of facial expression and present a method to analyze and visualize the expression subtlety by using dynamic high resolution data. We first introduced and compared facial expression databases used by computer vision researchers. Then a series of facial expression analysis techniques are discussed in terms of facial expression representation, facial tracking algorithms and expression classification frameworks. In addition, we briefly talked about algorithm evaluation issues for facial expression analysis systems. After that, we present a tool for the analysis and visualization of facial expression subtlety using dynamic 3D data and video data of high temporal frequency. Experiments demonstrate promising results on the subtlety analysis of facial expressions.

From the author's point of view, the following questions are still open for future research:

1. **Is facial expression analysis a recognition problem?** This question is related to the representation of facial expression. Currently, most of works either classify expressions into one of six universal emotion categories or detect the present of Action Units based on FACS description. Although researchers began to pay attention to the dynamic nature of facial expressions, these representations (basic emotions or FACS) are static in nature. A truly dynamic representation will be necessary to capture the density, asymmetry and personality of facial expressions.
2. **Does automatic analysis of facial expression really work?** Nowadays, most of researchers report recognition rate of more than 90 percent in the literature. However, this rate is either achieved on a small dataset or a deliberated expression database under controlled environment. Many factors contained in spontaneous

expressions such as illumination change, head motion, partial occlusion, speech related expression, could fail these systems. Furthermore, noisy measurements of images could significantly reduce recognition rate of these systems. It's time to go one step further to tackle these difficulties in spontaneous expression analysis.

3. **2D or 3D?** Automatic facial expression analysis from 2D images or video sequences has been studied for more than fifteen years while 3D expression analysis is still a newborn for the computer vision community. 3D expression data collection, 3D facial feature tracking and 3D representation of facial expressions remain open for all researchers in this area.

References

[1] M.S. Bartlett, B. Braathen, G.L. Littlewort-Ford, J. Hershey, J. Fasel, T. Mark, E. Smith, T.J. Sejnowski, and J.R. Movellan: Automatic Analysis of Spontaneous Facial Behavior: A Final Project Report, Technical Report MPLab-TR2001.08, Univ. of California at San Diego, Dec. 2001.

[2] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan, Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior, *In Proc. of IEEE CVPR* 2005

[3] M.S. Barlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan: Fully Automatic Facial Action Recognition in Spontaneous Behavior, *In Proc. of IEEE FGR* 2006

[4] J.N. Bassili, "Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Area of the Face," J. Personality and Social Psychology, vol. 37, pp. 2049-2059, 1979.

[5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, IEEE Trans. on PAMI, vol. 19, no. 7, pp. 711-720, July 1997

[6] M.J. Black and Y. Yacoob, "Recognizing Facial Expression in Image Sequences Using Local Parameterized Models of Image Motion," Int'l J. Computer Vision, vol. 25, no. 1, pp. 23-48, 1997.

[7] V. Blanz, T. Vetter, A morphable model for the synthesis of 3D faces. SIGGRAPH 1999

[8] J. Buhmann, J. Lange, and C. von der Malsburg, Distortion Invariant Object Recognition -- Matching Hierarchically Labelled Graphs, Proc. Int'l Joint Conf. Neural Networks, pp. 155-159, 1989.

[9] W. Cai, J. Wang, Adaptive multiresolution collocation methods for initial boundary value problems of nonlinear PDEs. SIAM Journal of Numerical Analysis, 33, 1996 937-970.

[10] J. Chai, J. Xiao and J. Hodgins: Vision-based Control of 3D Facial Animation, *In*

Proc. of ACM/Eurographics Symposium on Computer Animation, 2003.

[11] Y. Chang, C. Hu, R. Feris, M.Turk: Manifold based analysis of facial expression, Journal of Image and Vision Computing, 2005

[12] I. Cohen, N. Sebe, A. Garg, L.S. Chen, and T. Huang: Facial expression recognition from video sequences: temporal and static modeling, *Journal of Computer Vision and Image Understanding*, 2003

[13] J. Cohn, T. Kanade, T. Moriyama, Z. Ambadar, J. Xiao, J. Gao, and H. Imamura: A Comparative Study of Alternative FACS Coding Algorithms, Tech. Report CMU-RI-TR-02-06, Robotics Institute, Carnegie Mellon University, November, 2001.

[14] J.F. Cohn, A.J. Zlochower, J.J. Lien, and T. Kanade, Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression, Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 396-401, 1998.

[15] T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham, Active Shape Models -- Their Training and Application, Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38-59, Jan. 1995.

[16] G.W. Cottrell and J. Metcalfe, EMPATH: Fface, Emotion, Gender Recognition Using Holons,º Advances in Neural Information Processing Systems 3, R.P. Lippman, ed., pp. 564-571, 1991.

[17] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski: Classifying facial actions. *IEEE Trans. PAMI*, 21(10):974-989, 1999.

[18] F. Dornaika and F. Davoine: Simultaneous Facial Action Tracking and Expression Recognition Using a Particle Filter, ICCV 2005.

[19] G.J. Edwards, T.F. Cootes, and C.J. Taylor, Face Recognition Using Active Appearance Models, Proc. European Conf. Computer Vision, vol. 2, pp. 581-695, 1998.

[20] P. Ekman and W.V. Friesen, Facial Action Coding System (FACS): Manual. Palo Alto, Calif: Consulting Psychologists Press, 1978.

[21] P Ekman, W. V. Friesen and M. O'Sullivan: Smiles when lying, *Journal of Personality and Social Psychology* 1988

[22] P. Ekman: Strong Evidence for Universals in Facial Expressions: A Reply to

Russell's Mistaken Critique, Psychological Bulletin, vol. 115, no. 2, pp. 268-287, 1994.

[23] P. Ekman, E.L. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS),* Oxford University Press, 1997

[24] H. Ellgring: Nonverbal expression of psychological states in psychiatric patients. *European Archives of Psychiatry and Neurological Sciences* 1986

[25] I. Essa and A. Pentland, Coding, Analysis Interpretation, Recognition of Facial Expressions, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 757-763, July 1997.

[26] M. Frank, P. Ekman: The ability to detect deceit generalizes across different types of high-stake lies, *Journal of Personality & Social Psychology*, 1997.

[27] L. Gralewski, N. Campbell, I. Penton-Voak: Using a Tensor Framework for the Analysis of Facial Dynamics, IEEE FGR 2006

[28] H. Gu, Q. Ji: Facial event classification with task oriented Dynamic Bayesian Network, *In Proc. of IEEE CVPR* 2004.

[29] M. Heller and V. Haynal: Depression and Suicide Faces, *Cahiers Psychiatriques Genevois 1994*

[30] H. Hong, H. Neven, and C. von der Malsburg, Online Facial Expression Recognition Based on Personalized Galleries, Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 354-359, 1998.

[31] C.L. Huang and Y.M. Huang, Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification, J. Visual Comm. and Image Representation, vol. 8, no. 3, pp. 278-290, 1997.

[32] C.E. Izard, The maximally discriminative facial movement coding system (MAX), unpublished manuscript, available from Instructional Resource Center, University of Delaware.

[33] T. Kanade, J.Cohn, and Y. Tian: Comprehensive database for facial expression analysis, *In Proc FGR* 2000.

[34] A. Kanaujia and D. Metaxas: Recognizing Facial Expressions by Tracking Feature

Shapes, IEEE ICPR 06.

[35] G.D. Kearney and S. McKenzie, Machine Interpretation of Emotion: Design of Memory-Based Expert System for Interpreting Facial Expressions in Terms of Signaled Emotions (JANUS), Cognitive Science, vol. 17, no. 4, pp. 589-622, 1993.

[36] S. Kimura and M. Yachida, Facial Expression Recognition and Its Degree Estimation, In Proc. of IEEE CVPR, pp. 295-300, 1997.

[37] H. Kobayashi and F. Hara, Recognition of Six Basic Facial Expressions and Their Strength by Neural Network, Proc. Int'l Workshop Robot and Human Comm., 1992.

[38] H. Kobayashi and F. Hara, Facial Interaction between Animated 3D Face Robot and Human Beings, Proc. Int'l Conf. Systems, Man, Cybernetics,, pp. 3,732-3,737, 1997.

[39] Y. C. Lee, D. Terzopoulos, K. Waters. Realistic Face Modeling for Animation. SIGGRAPH 1995, pp. 55-62

[40] J.J. Lien, T. Kanade, J.F. Cohn, and C. Li: Detection, Tracking, and Classification of Action Units in Facial Expression, J. Robotics and Autonomous Systems, vol. 31, pp.131-146, 1997.

[41] B.D. Lucas, T. Kanade, An Iterative Image Registration Technique with an Application in Stereo Vision, Seventh International Joint Conference on Artificial Intelligence, 1981.

[42] M. Lyons, et al. Automatic Classification of Single Facial Images. *IEEE Trans. PAMI,* 21(12):1357-1362, 1999.

[43] M.J. Lyons, J. Budynek, and S. Akamatsu, Automatic Classification of Single Facial Images, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 12, pp. 1,357-1,362, 1999.

[44] A. M. Martinez and R. Benavente, *The AR Face Database*, CVC Technical Report #24, June 1998

[45] K. Mase, Recognition of Facial Expression from Optical Flow, IEICE Trans., vol. E74, no. 10, pp. 3,474-3,483, 1991.

[46] K. Matsuno, C.W. Lee, and S. Tsuji, Recognition of Facial Expression with Potential Net, Proc. Asian Conf. Computer Vision, pp. 504-507, 1993.

[47] K Matsuno, C. Lee, S. Kimura, S. Tsuji: Automatic Recognition of Human Facial Expressions, ICCV 95

[48] Y. Moses, D. Reynard, and A. Blake, Determining Facial Expressions in Real Time, Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 332-337, 1995.

[49] J. Y. Noh, U. Neumann, Expression cloning, SIGGRAPH 2001

[50] C. Padgett and G.W. Cottrell, Representing Face Images for Emotion Classification, Proc. Conf. Advances in Neural Information Processing Systems, pp. 894-900, 1996.

[51] M. Pantic and L.J.M. Rothkrantz, Expert System for Automatic Analysis of Facial Expression, Image and Vision Computing J., vol. 18, no. 11, pp. 881-905, 2000a

[52] M. Pantic and L. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424-1445, Dec. 2000b

[53] A. Pentland, B. Moghaddam, and T. Starner, View-Based and Modular Eigenspaces for Face Recognition, Proc. Computer Vision and Pattern Recognition, pp. 84-91, 1994.

[54] F. I. Parke, Computer Generated Animation of Faces. Proc. ACM annual conf., 1972

[55] R.W Picard, E. Vyzas, and J. Healey: Toward Machine Emotional Intelligence: Analysis of Affective Physiological State, *IEEE Trans on PAMI* 2001

[56] P. J. Phillips, Hyeonjoon Moon, S. A. Rizvi, and P. J. Rauss, *The FERET evaluation methodology for face recognition algorithm*, IEEE Trans. on PAMI, vol. 22, no. 10, pp. 1090-1104, October 2000

[57] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, *Overview of the face recognition grand challenge*, Proc. of CVPR05, no. 1, pp. 947–954, June 2005

[58] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, D. Salesin: Synthesizing Realistic Facial Expressions from Photographs. SIGGRAPH 1998: 75-84

[59] F. Pighin and J.P. Lewis, Performance-Driven Facial Animation, SIGGRAPH 2006

[60] A. Rahardja, A. Sowmya, and W.H. Wilson, A Neural Network Approach to

Component versus Holistic Recognition of Facial Expressions in Images, SPIE, Intelligent Robots and Computer Vision X: Algorithms and Techniques, vol. 1,607, pp. 62-70, 1991.

[61] M. Rosenblum, Y. Yacoob, and L. Davis, Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture, Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pp. 43-49, 1994.

[62] S. T. Roweis, L.K. Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding, Science, 2000

[63] J.A. Russell: Is There Universal Recognition of Emotion from Facial Expression?, Psychological Bulletin, vol. 115, no. 1, pp. 102-141, 1994.

[64] F. S. Samaria and A. C. Harter 'Parameterisation of a stochastic model for human face identification', *Proc. of 2nd IEEE workshop on Applications of Computer Vision*, pp. 138-142, 1994

[65] K. Schmidt, Z. Ambadar, J. Cohn and L.I. Reed: Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling, *Journal of Nonverbal Behavior*, 2006

[66] N. Sebe, M. Lew, I. Cohen, Y Sun, T. Gevers, and T. Huang: Authentic facial expression analysis, *IEEE FGR* 2004.

[67] E. Sifakis, I. Neverov, R. Fedkiw: Automatic Determination of Facial Muscle Activations from Sparse Motion Capture Marker Data, SIGGRAPH 2005

[68] E. Sifakis, A. Selle, A. Robinson-Mosher, R. Fedkiw, Simulating Speech with a Physics-Based Facial Muscle Model, ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA), 2006.

[69] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination and expression database. *IEEE Trans. PAMI*, 25(12), 2003

[70] E. Simoncelli, Distributed Representation and Analysis of Visual Motion, PhD thesis, Massachusetts Inst. of Technology, 1993.

[71] J. B. Tenenbaum, V. Silva, J.C. Langford: A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 2000

[72] Y. Tian, T. Kanade, and J.F. Cohn, “Recognizing Action Units for Facial Expression Analysis,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 97-115, Feb. 2001.

[73] Y. Tong, W. Liao, Q. Ji: Inferring Facial Action Units with Causal Relations, IEEE CVPR 2006.

[74] H. Ushida, T. Takagi, and T. Yamaguchi, Recognition of Facial Expressions Using Conceptual Fuzzy Sets, Proc. Conf. Fuzzy Systems, vol. 1, pp. 594-599, 1993

[75] P. Vanger, R. Honlinger, and H. Haken, Applications of Synergetics in Decoding Facial Expression of Emotion, Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 24-29, 1995.

[76] M. Vasilescu and D. Terzopoulos. Multilinear Analysis of Image Ensembles: Tensorfaces. In ECCV, 2002a.

[77] M. Vasilescu and D. Terzopoulos. Multilinear Image Analysis for Facial Recognition. In ICPR, 2002b.

[78] M.A.O. Vasilescu, D. Terzopoulos: TensorTextures: Multilinear Image-Based Rendering, *Proc. ACM SIGGRAPH* 2004

[79] P. Viola and M.J. Johns, Robust Real-Time Face Detection, IJCV 2004

[80] D. Vlasic, M. Brand, H. Pfister, J. Popović: Face Transfer with Multilinear Models ACM Transactions on Graphics 24(3), 2005, pages 426-433

[81] Y. Wang, X. Huang, C.S. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, P. Huang, High resolution acquisition, learning and transfer of dynamic 3-d facial expressions, *Computer Graphics Forum*, 2004

[82] M. Wang, Y. Iwai, and M. Yachida, Expression Recognition from Time-Sequential Facial Images by Use of Expression Change Model, Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 324-329, 1998.

[83] K. Waters. A Muscle Model for Animating Three-dimensional Facial Expression, SIGGRAPH 1987, vol. 21 pp. 17-24

[84] K. Waters, S. Terzopoulos, Modeling and Animating Faces using Scanned Data, Journal of Visualization and Computer Animation, 1991, Vol. 2, No. 4, pp. 123–128

[85] L. Williams, Performance-driven facial animation. In Proceedings of SIGGRAPH 1990

[86] Y. Yacoob and L. Davis, Recognizing Facial Expressions by Spatio-Temporal Analysis, Proc. Int'l Conf. Pattern Recognition, vol. 1, pp. 747-749, 1994.

[87] Y. Yacoob and L. Davis, Recognizing Human Facial Expressions From Long Image Sequences Using Optical Flow, PAMI 1996

[88] L. Yin, X. Wei, Y. Sun, J. Wang, M.J. Rosato: A 3D Facial Expression Database For Facial Behavior Research, *IEEE FGR* 2006

[89] M. Yoneyama, Y. Iwano, A. Ohtake, and K. Shirai, Facial Expressions Recognition Using Discrete Hopfield Neural Networks, Proc. Int'l Conf. Information Processing, vol. 3, pp. 117-120, 1997.

[90] Y. Zhang, Q. Ji: Analysis and Synthesis of Facial Image Sequences Using Physical and anatomical models, *IEEE Trans on. PAMI* 2005

[91] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, Comparison between Geometry-Based and Gabor Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron, Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 454-459, 1998.

[92] Li Zhang, Noah Snavely, Brian Curless, and Steve Seitz, 'Spacetime Faces: High-resolution capture for modeling and animation,' *Proc. of ACM SIGGRAPH 2004*

[93] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld: Face Recognition: a Literature Survey. *ACM Computing Surveys*, 35(4), 2003.

[94] J. Zhao and G. Kearney, Classifying Facial Emotions by Backpropagation Neural Networks with Fuzzy Inputs, Proc. Conf. Neural Information Processing, vol. 1, pp. 454-457, 1996.

[95] M. J. Black and P. Anandan, The Robust Estimation of Multiple Motions: Parametric and Piecewise-smooth Flow Fields, Computer Vision and Image Understanding, CVIU, 63(1), pp. 75-104, Jan. 1996.

[96] Brian Cabral and Leith (Casey) Leedon, Imaging Vector Fields Using Line Integral Convolution. Proc. of ACM SIGGRAPH 1993 pp. 263-272, 1993

[97] Yang Wang, Mohit Gupta, Song Zhang, Sen Wang, Xianfeng Gu, Dimitris Samaras, Peisen Huang. High Resolution Tracking of Non-Rigid 3D Motion of Densely Sampled Data Using Harmonic Maps, In Proc. of ICCV, pp. I: 388-395, 2005.

[98] Jarke J. van Wijk, Image Based Flow Visualization. ACM Transactions on Graphics, special issue, Proceedings ACM SIGGRAPH 2002