การตรวจจับอารมณ์จากใบหน้าในชุดลำดับภาพโดยใช้การติดตามจุดบนใบหน้าด้วยการประมาณค่าโมเดลความน่าจะเป็นด้วยอนุภาค

นายธาดา จิระจรัส

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาคณิตศาสตร์ประยุกต์และวิทยาการคณนา

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2557

EMOTION DETECTION FROM FACES IN IMAGE SEQUENCES USING FACIAL POINT TRACKING WITH PROBABILISTIC MODEL ESTIMATION WITH PARTICLES

Mr. Thada Jirajaras

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Applied Mathematics and Computational Science
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2014

Thesis Title	EMOTION DETECTION FROM FACES IN IMAGE SEQUENCES USING FACIAL POINT TRACKING WITH PROBABILISTIC MODEL ESTIMATION WITH PARTICLES
By	Mr. Thada Jirajaras
Field of Study	Applied Mathematics and Computational Science
Thesis Advisor	Assistant Professor Rajalida Lipikorn , Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

Deputy Dean, Acting Dean,
..... The Faculty of Science
(Associate Professor Somkiat Ngamprasertsith, Dr.de l'INPT)

THESIS COMMITTEE

.....Chairman
(Assistant Professor Krung Sinapiromsaran, Ph.D.)

.....Thesis Advisor
(Assistant Professor Rajalida Lipikorn , Ph.D)

.....Examiner
(Assistant Professor Nagul Cooharojananone, Ph.D.)

.....External Examiner
(Associate Professor Zeng Lertmanorat, Ph.D.)

ธาดา จิระจรัส : การตรวจจับอารมณ์จากใบหน้าในชุดลำดับภาพโดยใช้การติดตามจุดบนใบหน้าด้วยการประมาณค่าโมเดลความน่าจะเป็นด้วยอนุภาค.
(EMOTION DETECTION FROM FACES IN IMAGE SEQUENCES USING FACIAL POINT TRACKING WITH PROBABILISTIC MODEL ESTIMATION WITH PARTICLES)
อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ.ดร. รัชลิดา ลิปิกรณ์, 50 หน้า.

การตรวจจับอารมณ์มีความเกี่ยวข้องในสาขาต่างๆ อาทิเช่น ด้านการศึกษาพฤติกรรม, การกายภาพบำบัด และการเรียนรู้กับคอมพิวเตอร์ เป็นต้น ถ้าคอมพิวเตอร์สามารถตรวจจับอารมณ์ได้ จะทำให้คอมพิวเตอร์เข้ามามีบทบาทสำคัญในสาขาต่างๆเหล่านี้ได้ กระบวนการตรวจจับอารมณ์ประกอบไปด้วย การตรวจหาใบหน้า การสกัดลักษณะเฉพาะที่ได้จากใบหน้า และการใช้สิ่งที่สกัดได้มาทำการแยกแยะอารมณ์ต่างๆออกจากกัน จุดบนใบหน้าถูกติดตามโดยใช้เพียงข้อมูลตำแหน่งของจุดบนใบหน้าจากเฟรมก่อนหน้าและข้อมูลรายละเอียดสีรอบๆ จุดบนใบหน้าของเฟรมที่แสดงหน้าอารมณ์เป็นกลาง ข้อมูลจากหน้าอารมณ์เป็นกลางทำให้การติดตามจุดมีความแม่นยำมากยิ่งขึ้น อีกทั้งข้อมูลตำแหน่งของจุดจากเฟรมก่อนหน้าเข้ามามีส่วนช่วยอย่างมากเพราะว่าจุดในเฟรมปัจจุบันมีแนวโน้มที่จะวางตัวอยู่ใกล้เคียงกับบริเวณของจุดที่อยู่บนเฟรมก่อนหน้า โดยการติดตามจุดจะใช้โมเดลทางความน่าจะเป็นโดยใช้การประมาณค่าการคำนวนหาค่าความคาดหวังของตำแหน่งจุดบนใบหน้าด้วยอนุภาค เมื่อการติดตามจุดมาถึงเฟรมสุดท้ายซึ่งมีการแสดงออกทางอารมณ์สูงสุด เราจะสกัดลักษณะที่ได้จากเฟรมสุดท้ายนี้เพื่อมาทำการแยกแยะอารมณ์ จุดเด่นของการใช้อนุภาคมาประมาณค่าความน่าจะเป็นคือการทำให้การหาค่าความคาดหวังของโมเดลความน่าจะเป็นมีความซับซ้อนต่ำ

ภาควิชา	คณิตศาสตร์ และวิทยาการคอมพิวเตอร์	ลายมือชื่อนิสิต	.....
		ลายมือชื่อ อ.ที่ปรึกษาหลัก	.....
สาขาวิชา	คณิตศาสตร์ประยุกต์ และวิทยาการคณนา		
ปีการศึกษา	2557		

5572197723 : MAJOR APPLIED MATHEMATICS AND COMPUTATIONAL SCIENCE KEYWORDS : PROBABILITY / PARTICLE ESTIMATION

THADA JIRAJARAS : EMOTION DETECTION FROM FACES IN IMAGE SEQUENCES USING FACIAL POINT TRACKING WITH PROBABILISTIC MODEL ESTIMATION WITH PARTICLES

ADVISOR : ASST. PROF. RAJALIDA LIPIKORN, Ph.D., 50pp.

Emotion detection is related to many fields such as behavioral study, rehabilitation, e-learning, etc. If computers can detect emotions, they will be able to play an important role in applications of these fields. Emotion detection process includes face detection, features extraction, and, emotion classification. Facial points are tracked using only the spacial information from the previous frame and texture information from the first frame. Texture information from the neutral face helps the tracking procedure to have accurate facial feature localization in each frame. The feature locations from the previous frame are used to predict the feature locations of the current frame. Location information from the previous frame helps the tracking procedure to track the facial feature locations of the current frame easily because the current feature locations tend to be located near the feature locations of the previous frame. Our expectation is to classify an emotion from the last frame (peak of emotion) in an image sequence. We use textures form the neutral face and the facial points from the previous frame to form a probabilistic model. After that, the facial points in each frame are assigned using the particle estimation to find the expected values of facial point locations. Then, we extract emotion from features produced by these points by using a classification. The benefit of using particles for probabilistic estimation is that finding expectation value of probabilistic model has low complexity.

Department : Mathematics and Computer Science Student's Signature .....
Advisor's Signature .....

Field of Study : Applied Mathematics and Computational Science

Academic Year : 2014

ACKNOWLEDGEMENTS

I would like to express my thanks to my advisor, Assistant Professor Dr. Rajalida Lipikorn, for her invaluable help and encouragement throughout the course of this study. I am most grateful for her suggestions and teaching, not only research methodologies but also many methodologies in life. Without her support, this thesis could not have been completed.

Additionally, I would like to thank those whose names are not mentioned here but greatly inspired and encouraged us until this thesis comes to the end.

Finally, I most gratefully acknowledge my parents and my friends for all their support throughout the period of this study. I also most gratefully thank to the Applied Mathematics and Computational Science, Faculty of Science and Graduate school, Chulalongkorn University for financial support to the national conference.

CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Page

Page

CHAPTER I

Introduction

1.1 Rationale for This Thesis

Emotion detection is related to many fields such as behavioral study, rehabilitation, e-learning, etc. If computers can detect emotions, they will be an important tool in applications of these fields as well. Emotion detection from human faces is one of non-invasive methods to detect emotions from facial expression. Emotion detection from faces is still an active research in computer vision. It is easy for humans to detect emotion expression from faces. However, this is not an easy task for computers. There are many attempts to accomplish this job. The overview of emotion detection methods in most papers are similar. Firstly, a face is detected. Secondly, features from a face are extracted. Finally, emotions are classified using these features.

The goal of our work is to find good features from a face to classify emotions. Good features in our work should have three properties. Firstly, features should be extracted from any single frame in an image sequence using the templates from the neutral face frame only. Texture information from the neutral face helps the tracking procedure to have accurate facial feature localization. This makes the emotion detection system using this algorithm can be used with a new face sequence whose only neutral face information has been added to the algorithm. Secondly, features should be extracted from any single frame in an image sequence using only the location information from the previous frame. Location information from the previous frame helps the tracking procedure to track the facial feature locations of the current frame easily because the current feature locations tend to be located near the feature locations of the previous frame.

Emotion detection is useful for computer to analyze, collect, or response to

the emotion information. Our proposed work focuses on six basic emotion classification: happiness, sadness, anger, surprise, disgust, and fear. The ability of detecting emotion can be an important step to human-computer interaction. The difficulty of this area is to find and interpret important features from human faces that vary from one to another and from illumination. If computers can immitate human better, they will have the ability to help human in more areas than they have been able to do.

1.2 Statement of The Problem

Our expectation is to classify an emotion from the last frame (peak of emotion) in an image sequence from tracked facial points using only texture template from the first frame (neutral emotion) and spacial information from the previous frame because in the real would application that detects faces from a video sequence, sometimes there might not be any texture information from other emotional frames except the neutral frame to train the system. We avoid the tracking methods that need the training process thus we choose to adapt the probabilistic model together with the spatial and texture models in our work.

1.3 Objectives

The objectives of this work is to be able to classify an emotion from an image sequence labeled as one of six basic emotions.

1.4 Method

Our study research focuses on finding a method that can classify emotions from 2D gray-scale image sequences using the image processing, the probabilistic model and the classification technique.

We use textures from the neutral face and the facial points from the previous frame to form a probabilistic model for facial point tracking. Textures from the neutral face are used to form an observation or texture model. The facial points

from the previous frame are used to form a transition or spatial model. After that, the facial points in each frame are assigned. Then, we extract emotion from features produced by these points by using a classification technique.

1.5 Structure of The Thesis

In this paper, we start with the literature review and preliminary in chapter 2 and go forward to our proposed method in chapter 3. Experimental results and discussion are presented in chapter 4. In the literature review chapter and preliminary, we introduce emotion expression, image processing, basic mathematics that are correlated to our work, and related papers. In the proposed-method chapter, we introduce our probability model for facial point tracking with probabilistic estimation by particles and classification technique for classifying an emotion. In chapter 4, we show the accuracy rate of facial point tracking and emotion classification as well as discussion of strong points and limitations for this work.

CHAPTER II

Literature Review and Preliminary

2.1 Literature Review

Emotion detection can give some information that is correlated with many applications such as behavioral studying, rehabilitation, or e-learning in such the way that computer can analyze, collect, or respond to such information.

Works in this area have tried to make computers be able to understand human better so that they can help human in more areas than they have been able to do. One benefit of extracting interesting features from images is that this method does not have any invasive process and sensor attachment. The ability of detecting emotion is one of several important steps in human-computer interaction. The difficulty of this area is to find and interpret important features that vary from one to another and from illumination. Normally, there are challenges in developing a system for emotion detection in many steps such as facial point detection [1, 2], facial point tracking [3], and emotion classification [4, 5] because all of them need to use information from the first or the previous frame to complete their task.

C. Bouvier[6] proposed the method for lip segmentation that does not require information from other frames but they still need to use color information from RGB images. Some researches use edge detection [7] or blob finding [8] to reduce the possible area of facial feature positions. Many researches use image sequences of Cohn-Kanade(CK) [9] which have one of six basic emotions labeled. The description of the database can be seen in Table 2.1 and Table 2.2. Saeed, A. and Al-Hamadi, A. and Niese, R. and Elzobi, M. suggest effective geometric features [5] from chosen facial points. This paper shows that choosing even a few but good features for a classifier can lead to the good results. SVM is the classifier that is used in this paper.

Properties	Descriptions
Number of subjects	18 to 50 years of age, 69% female, 81%, Euro-American, 13% Afro-American, and 6% other groups
Gray/Color	Eight-bit gray
Resolution	640 x 490
Frame rate	12 frame/sec

Table 2.1: Description of Cohn-Kanade AU-Coded facial expression database

label	Number of image sequences
Angry	45
Contempt	18
Disgust	59
Fear	25
Happy	69
Sadness	28
Surprise	83

Table 2.2: Number of image sequences that are labeled as one of six basic emotions

Our expectation is to detect an emotion from any frame in an image sequence with facial point tracking using just information from the previous frame or the first frame, which contains the neutral face. In real world applications when detecting emotion from a video camera, sometimes there might be only information from the neutral or peak emotional face. In these cases, using large information to train the facial point tracking model is impossible, thus we try to detect features from faces using as less information as possible to train the tracking and classification model.

2.2 Preliminary

2.2.1 Probabilistic Model

Probabilistic model is a statistical method. In our case, we use a transition model and an observation model to form the probabilistic model for facial point tracking. In this preliminary part, a simple example of the transition model and the observation model are shown in Eq. (2.1) and (2.2). $\mathbf{X_k} = (x^{(k)}, y^{(k)})$ are coordinates for the horizontal and vertical axes of the current frame k, where x is the horizontal coordinate and y is vertical coordinate. $\mathbf{X_{k-1}} = (x^{(k-1)}, y^{(k-1)})$ are x and y coordinates of the previous frame $k-1$.

Transition Model

$$\mathbf{X_k} = \begin{bmatrix} Triangle(left_x, right_x, mode = x^{(k-1)}) \\ Triangle(left_y, right_y, mode = y^{(k-1)}) \end{bmatrix} \tag{2.1}$$

- $\mathbf{X_k}$ is the facial point coordinates (x, y) in the current frame.
- $\mathbf{X}_{k-1}$ is the facial point coordinates in the previous frame.
- $Triangle(left, right, mode)$ is the triangular probability distribution where $left, right, mode$ are left bound, right bound and mode(highest peak) of this distribution, respectively.

The meaning of this Eq. (2.1) is that the new facial point positions depend on and stay close to the facial point positions from the previous frame. The triangular distribution of $x^{(k)}$ in Eq. (2.1) is shown in Figure 2.1 and it is similar for $y^{(k)}$.

Observation Model

The observation model uses the texture information from the neutral face as the template. Normalized cross-correlation values $\rho(\mathbf{c}, \mathbf{z})$ between the template $\mathbf{c}$ and the window $\mathbf{z}$ centered at the interested coordinates are used to form the probabilistic model. As seen in Figure 2.2, an example of the template of right eye corner is shown. The template is the window that contains pixel values centered

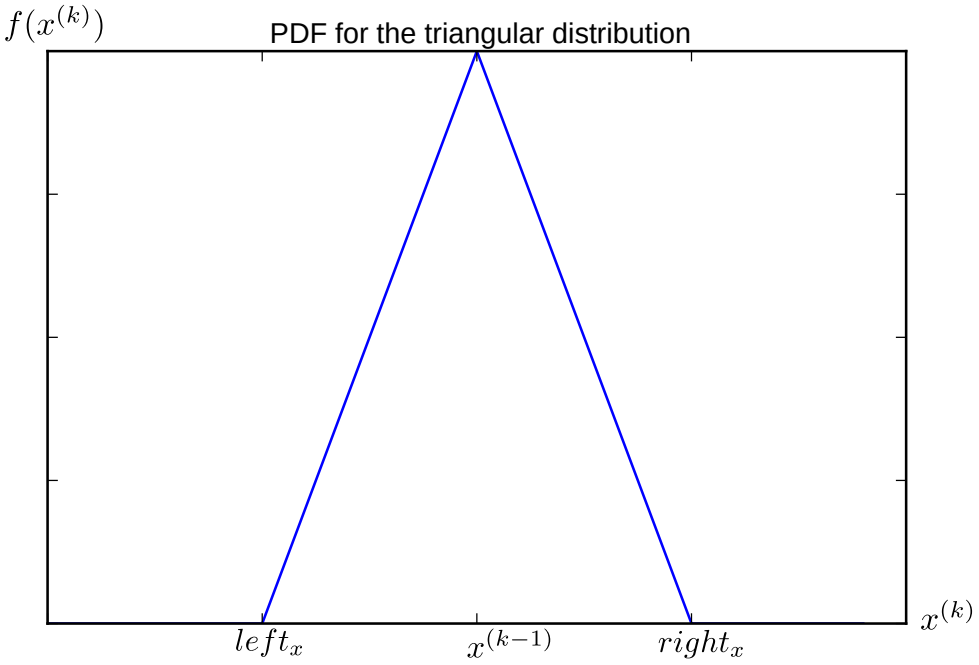

Figure 2.1: Triangular distribution: x-axis is the x coordinate of the frame k called $x^{(k)}$ and y -axis is the PDF for the triangular distribution $f(x^{(k)})$ of the random variable $x^{(k)}$

at the right eye corner point. Templates of other facial points to be tracked are defined similar to the right eye corner case.

$$p(\mathbf{z}|\mathbf{X}_k) = \Delta \times e^{\frac{(\rho(\mathbf{c},\mathbf{z})-1)}{0.05}} \tag{2.2}$$

where

$$\rho(\mathbf{c},\mathbf{z}) = \left| \frac{\sum_{r\in R^2}[c(r)-\bar{c}][z(r)-\bar{z}]}{\sqrt{\sum_{r\in R^2}[c(r)-\bar{c}]^2}\sqrt{\sum_{r\in R^2}[z(r)-\bar{z}]^2}} \right|, 0 \leq \rho \leq 1$$

- $\mathbf{c}$ conveys gray scale template centered at the facial point of the neutral face
- $c(r)$ conveys gray scale value at r^{th} pixel of template c
- $\mathbf{z}$ proposes gray scale window centered at coordinates $\mathbf{X}$

  where size of window can be chosen depending on how much texture around the coordinates $\mathbf{X}$ is needed.
- $z(r)$ denotes gray scale value of r^{th} pixel of window $\mathbf{z}$
- $\bar{c}$ is the average value of gray scale template $\mathbf{c}$
- $\bar{z}$ is the average value of $\mathbf{z}$
- $\mathbf{X}_k$ is the interested coordinates in current frame k
- Δ is the value for probabilistic normalization.

The meaning of this model is that the new facial point position probability distribution will be adapted using Eq. (2.2) by the observation or template $\mathbf{c}$ of the facial points from the neutral face. This model will reduce the probabilistic value at position with low correlation between texture $\mathbf{z}$ of that position $\mathbf{X}$ and the template $\mathbf{c}$.

2.2.2 Particle Estimation

In the complex probabilistic model, calculating joint distribution has the high cost of computing. Normally, in the case of two dependent random variables with

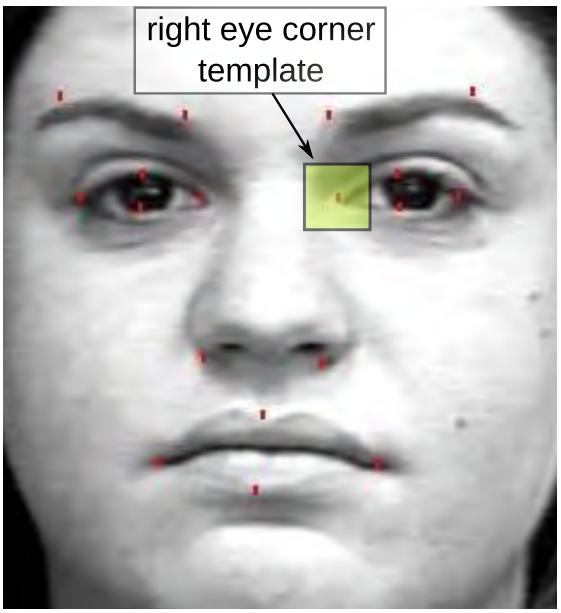

Figure 2.2: An example of the right eye corner template is shown. Small rectangular window centered at the right eye corner coordinates is the template **c**.

the same discrete domain of size n, the complexity of calculating over all joint domains is $O(n^2)$. However, in some real time applications, the estimation of joint distribution is evaluated by sampling m particles using the initial distribution of first random variable. Then it uses conditional distributions that distribute the probability of the first random variable to another random variable. With this method, the complexity of finding joint distribution can be reduced to $O(m)$. The simple example of using 100 particles to estimate the normal distribution (Figure 2.3) can be seen in Figure 2.4 and after rounding the domain as seen in Figure 2.5. In case of 100 particles, each particle equals 0.01 probability value because the summation of 100 probabilistic values of all particles needs to be 1. We can call that each particle has its weight equal 0.01.

In case of truncated normal distribution, we suggest two ways of using 100 particles, whose weights are 0.01, to estimate the probability distribution. Assume that our interested truncated normal distribution is as Figure 2.6. The first way is to sample 100 particles using the normal distribution as seen in Figure 2.3 then eliminate the particles that are not in range of -2 to 2. After elimination process,

we need to divide all particle weights with the summation of overall particle weights to make the summation of weights equals 1 again. Then we get the approximation of truncated normal distribution from the histogram of these particles. In order to make the number of particles to be 100 again, we need to use this approximation of truncated normal distribution for sampling the new 100 particles. Finally, the approximation of truncated normal distribution with 100 particle estimation with equal weight (= 0.01) for each particle is produced successfully. The second way is to generate 100 particles directly from the truncated normal distribution. The second one is better in term of accuracy because the first case has the particle elimination step that causes the estimation to be more rough because the number of particles are reduced to be less than 100 in between step. However, in some model with more than one random variable, the first way is better in term of speed. For example, the value of each random variable may be generated parallelly in a computer that has more than one processor and after that the particles that are not inside the bounding area can be eliminated after the sampling process. However, we use the second way to generate the particles in our work because it is more reliable.

2.2.3 Classification and Feature Selection

Support Vector Machine (SVM)

Support vector machine (SVM) is a classification technique that is reported to be the effective tool for classification and the complexity depends on the number of training samples but not on the dimension of features. The idea of this method is to map the features by a kernel to have higher dimensions but easier to be classified by linear discrimination. After mapping, the position of linear discrimination plane is optimized by maximizing the margin from support vector to discriminative line and minimizing sum of the square error $\sum_{i=1}^{n} \xi_i^2$ of classification as seen in Eq. (2.5). $\frac{Margin}{2}$ is the length from support vector to discriminative line. Discriminative line is defined by Eq. (2.3) where x is features and $\phi(x) \odot \omega$ is the dot product between mapped features $\phi(x)$ and parameter vector ω. The rela-

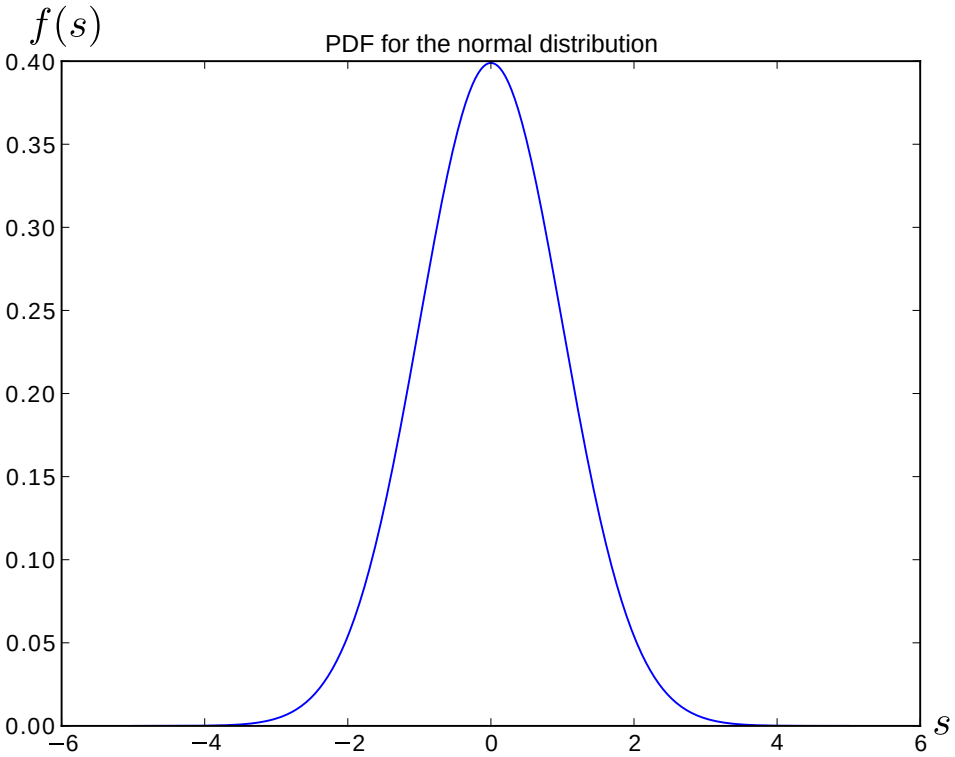

Figure 2.3: Normal distribution with $sd = 1$: $f(s)$ is the probability density function of the random variable s.

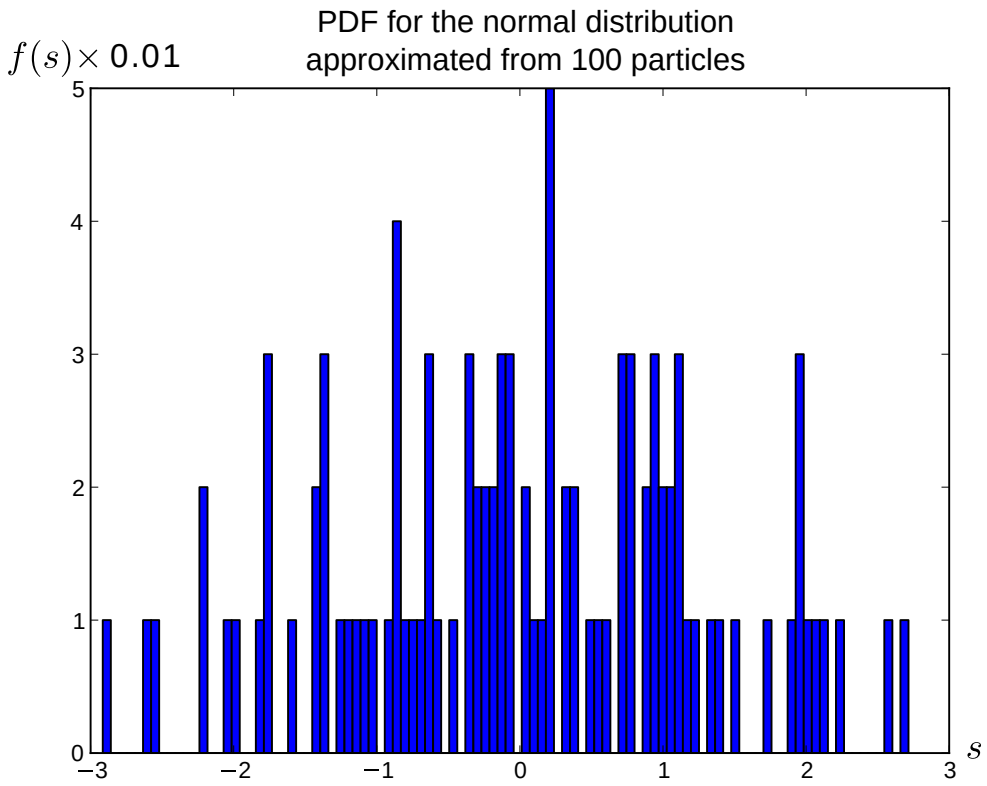

Figure 2.4: Normal distribution approximation from 100 particles with $sd = 1$: $f(s)\times 0.01$ is the probability density function of the random variable s.

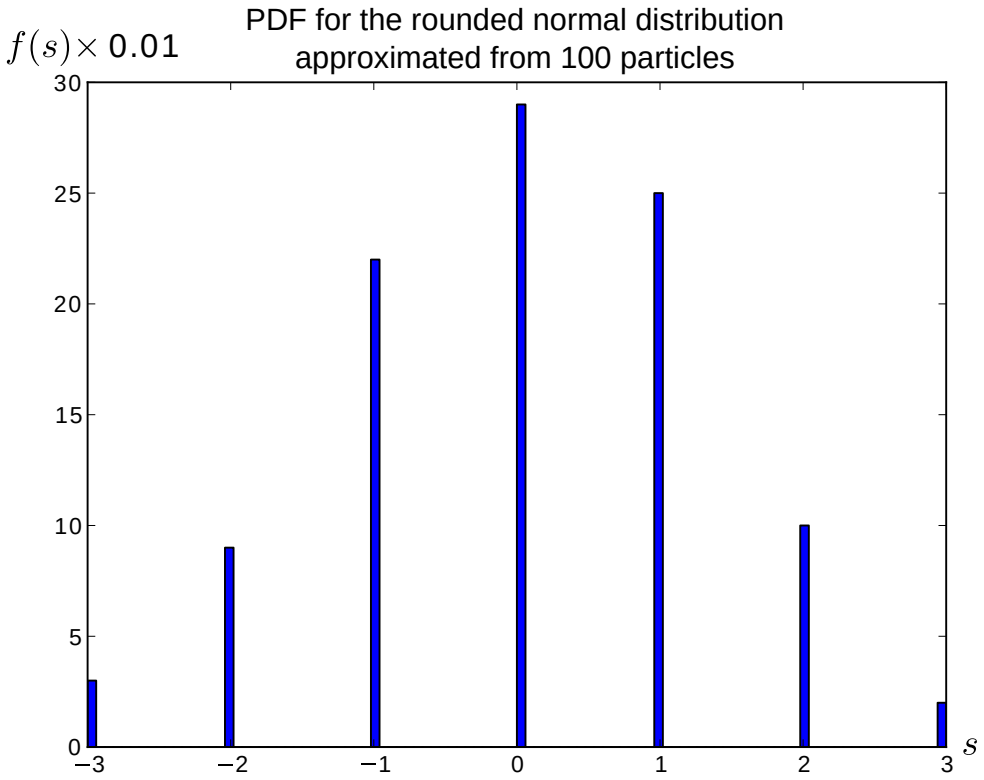

Figure 2.5: Rounded normal distribution approximation from 100 particles with $sd = 1$: $f(s)\times 0.01$ is the probability density function of the random variable s.

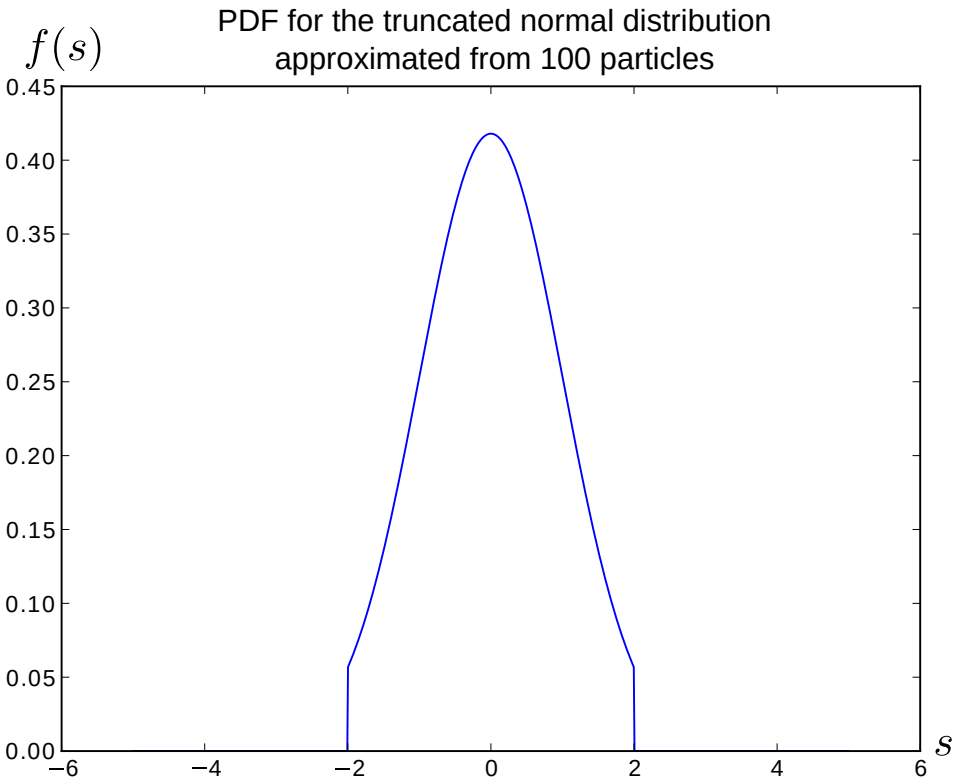

Figure 2.6: Normal distribution with $sd = 1$ truncated as $s - 2sd$ and $s + 2sd$: $f(s)$ is the probability density function of the random variable s.

tion between $Margin$ and ω is as Eq. (2.4). This method can only separate two groups apart from each other. In case of more than two groups to be separated, one-vs-one or one-vs-the rest techniques are used to solve the problem.

$$\phi(x) \odot \omega + b = 0 \tag{2.3}$$

$$Margin = \frac{2}{\|\omega\|} \tag{2.4}$$

$$\min \left(\|\omega\|^2 + c \sum_{i=1}^{n} \xi_i^2 \right) \tag{2.5}$$

N-fold Cross-Validation

In a classification problem, when one needs to test the performance of the classifier, n-fold cross-validation is an appropriate method for this purpose. The first step of the n-fold cross-validation is to partition a sample of data into n subsets. The cross-validation performs n rounds to reduce variability. One round of cross-validation uses one subset as the testing dataset and other $n-1$ subsets as the training dataset. The validation results are average accuracy over n rounds.

Relief (Feature Selection)

Relief is a fast feature selection method that requires only linear time in the number of given features and training instances. However, it is not suitable for discriminating between redundant features, and low numbers of training instances can cause a low accuracy of algorithm.

Assume that a dataset has n instances with p features, belonging to two known classes and each feature should be scaled to the interval [0 1]. The closest same-class instance is called 'near-hit', and the closest different-class instance is called 'near-miss'. The algorithm updates the weight W_i of each feature at iteration i by Eq. (2.6). x_i is the i^{th} sample from the dataset.

$$W_i = W_{i-1} - (x_i - \text{nearHit}_i)^2 + (x_i - \text{nearMiss}_i)^2 \tag{2.6}$$

Thus the weight of any given feature decreases if it differs from that feature in nearby instances of the same class more than nearby instances of the other class,

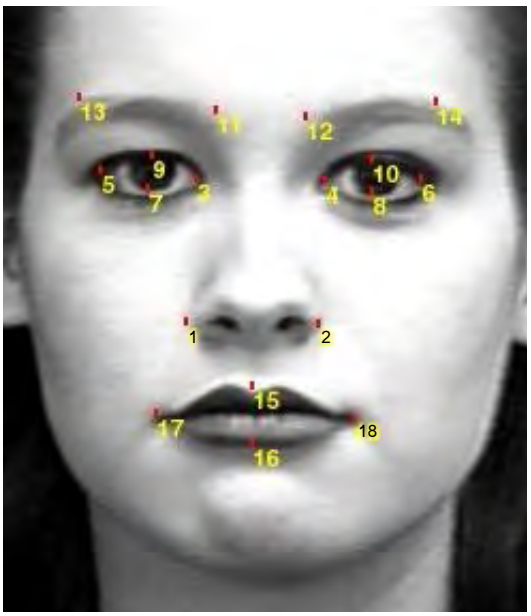

Figure 2.7: Facial points

and increases in the reverse case. the feature is selected if the its normalized weight is greater than a given threshold.

2.2.4 Facial Geometric Features

Facial Points around Eyes and Eyebrows

We use facial geometric features similar to ones used in [3] that use particle filter as the tracking procedure. Localizations of facial points in their research are done successively. The probability in each facial point location is estimated by particles. Facial points and geometric features EW, EMW, and HW can be seen in Figure 2.7 and Figure 2.8. Firstly, two points of nostrils 1 and 2 are localized. Secondly, i^{th} particle of points 3,4,5,and 6 are localized using the conditional equations below.

$$\frac{2}{3} \times EW \leq Width(x^i(3) - x^i(5)) \leq \frac{4}{3} \times EW \tag{2.7}$$

$$\frac{2}{3} \times EW \leq Width(x^i(6) - x^i(4)) \leq \frac{4}{3} \times EW \tag{2.8}$$

$$Height(x^i(3) - x^i(5)) \leq \frac{1}{3} \times HW \tag{2.9}$$

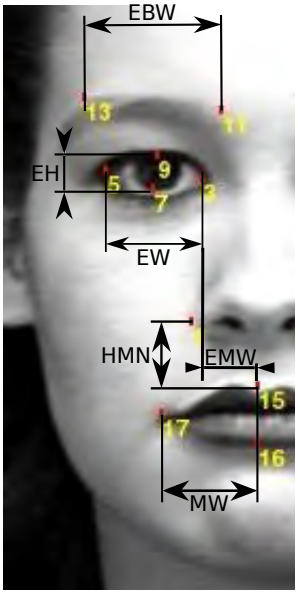

Figure 2.8: Geometric features

$$Height(x^i(6) - x^i(4)) \leq \frac{1}{3} \times HW \tag{2.10}$$

$$-\frac{1}{2} \times EMW \leq Width\Big(\frac{x^i(3) + x^i(4)}{2} - \frac{x(1) + x(2)}{2}\Big) \leq \frac{1}{2} \times EMW \tag{2.11}$$

$$-\frac{1}{2} \times EMW \leq Width\Big(\frac{x^i(5) + x^i(6)}{2} - \frac{x(1) + x(2)}{2}\Big) \leq \frac{1}{2} \times EMW \tag{2.12}$$

where $Width(x)$ denotes first component(x-axis) of x and $Height(x)$ denotes second component(y-axis) of x. And $x(s) = \frac{1}{N_s}\sum_{i=1}^{N_s} x^i(s)$ where $x^i(s)$ is coordinates of i^{th} particle dependent on facial point s and N_s is the total number of particles. Secondly, point 7 is assigned using conditional equations below. Similar equations are used in case of point 8.

$$W\Big(\frac{x(3) + x(5)}{2}\Big) - \frac{1}{4} \times EW \leq Width\Big(x^i(7)\Big) \leq W\Big(\frac{x(3) + x(5)}{2}\Big) + \frac{1}{4} \times EW \tag{2.13}$$

$$Height\Big(\frac{x(3) + x(5)}{2}\Big) \leq Height\Big(x^i(7)\Big) \leq Height\Big(\frac{x(3) + x(5)}{2}\Big) + EH \tag{2.14}$$

Eyebrow points 11 and 13 use the condition in Eq. (2.15) below and it is similar to points 12 and 14. The geometric feature EBW can be seen in Figure 2.8.

$$\frac{2}{3} \times EBW \leq Width\Big(x^i(11) - x^i(13)\Big) \leq \frac{4}{3} \times EBW \tag{2.15}$$

After getting accurate positions of points 1 to 8 and 11 to 14, point 9 is localized using equations (2.16) and (2.17). And it is similar to point 10.

$$W\Big(\frac{x(3)+x(5)}{2}\Big) - \frac{1}{4} \times EW \leq Width\Big(x^i(9)\Big) \leq W\Big(\frac{x(3)+x(5)}{2}\Big) + \frac{1}{4} \times EW \tag{2.16}$$

$$Height(x(11)) \leq Height(x^i(9)) \leq Height(x(7)) \tag{2.17}$$

Facial Points around Mouth

There are four interested points around the mouth. Nostril points are used as reference points for localizing accurately these four points. Top corner of the mouth, point 15, is localized using Eq. (2.18) and (2.19). The geometric features HMN and MW can be seen in Figure 2.8.

$$Width(x(1)) \leq Width(x^i(15)) \leq Width(x(2)) \tag{2.18}$$

$$Height\Big(x^i(15)\Big) \geq Height\Big(\frac{x(1)+x(2)}{2}\Big) + \frac{1}{2} \times HMN \tag{2.19}$$

After point 15 is assigned. The rest of the points are localized using equations below.

$$Width(x(1)) \leq Width(x^i(16)) \leq Width(x(2)) \tag{2.20}$$

$$Height(x^i(16) \geq Height(x(15)) \tag{2.21}$$

$$Width\Big(x(15)\Big) - 2 \times MW \leq Width\Big(x^i(17)\Big) \leq Width\Big(x(15)\Big) - \frac{1}{2} \times MW \tag{2.22}$$

$$Width\Big(x(15)\Big) + \frac{1}{2} \times MW \leq Width\Big(x^i(18)\Big) \leq Width\Big(x(15)\Big) + 2 \times MW \tag{2.23}$$

$$Height(x^i(17)) \geq Height(x(15)) \tag{2.24}$$

$$Height(x^i(18)) \geq Height(x(15)) \tag{2.25}$$

Figure 2.9: Binarization and *blob* finding

2.2.5 Adaptive Thresholding and Blob Finding

Adaptive thresholding is an adaptive method of binarization. We use the method that requires blocksize. A moving average of pixels over each block is called localMean of that window block. The threshold for each window block can be calculated as Eq. (2.26) where p is the constant that can be defined by a user, e.g. $p = 5$.

$$threshold = localMean - p \tag{2.26}$$

Blob finding looks for continuous light regions and return them as Blob features. Figures 2.9 shows adaptive binarization and finding *blob*. In this Figure, there are many *blob*s founded in *blob* finding process and one example of them is painted in green colour.

The interested *blob* can be used in the facial point tracking process to reduce the possible area of particles used in the probabilistic model.

CHAPTER III

Methodologies

We use a probabilistic model with particle estimation to assign the positions of facial points in each frame and extract the features from these points with an emotional label for training and testing the classifier. We use texture information from the first frame that is neutral face and spatial information from previous frame to form a probabilistic model that can indicate the correct position of the facial points. Then, these positions are processed to use with the classifier. Last output of this method for an image sequence is an assigned emotion that is one of the six basic emotions.

We propose the facial point tracking method similar to fiducial facial point tracking using particle filter and geometric features [3] by Fazli, S. and Afrouzian, R. and Seyedarabi, H. We improve the sampling state of particles of this paper by not canceling the particles in the impossible areas after sampling like the method in this paper but try to begin with sampling particles into the possible area as much as possible to avoid loosing particles that means loosing accuracy of probability estimation. The number of particles means how much the probability is approximated. The fewer particles, the probability distribution is more approximated. So we avoid the elimination or reducing the number of particles. Some researches use edge detection [7] or blob finding [8] to reduce the possible area of facial feature positions. In our work, we also use an image processing called local binarization to help assign some facial points. We use Cohn-Kanade (CK) dataset in our work [9]. We use image sequences of CK which have one of six basic emotions labeled. And we use the facial points of the first frame in each image sequence from text file containing the facial point positions for extracting texture around those interested facial points. The facial points from this database are assigned from the Active Appearance Model (AMM) method. In our work,

we choose only image sequences with given correct facial point positions in the first frame. Saeed, A. and Al-Hamadi, A. and Niese, R. and Elzobi, M. suggest effective geometric features [5] from chosen facial points. This paper shows that choosing even a few but good features for a classifier can lead to the good results. SVM is the classifier that is used in this paper.

3.1 Overview

Fig 3.1 shows the overview of emotion detection processes for each image sequence. Firstly, a frame is queried from an image sequences, If it is the first frame, we can get the existed facial point positions from Cohn-Kanade database. Then, these facial points are kept to use in the next frame to find current facial point positions. If it is not the first frame, it goes to facial point tracking process. In this process, previous positions of facial points are used to guide the locations of the current facial points. Finally, if this is the last frame, the tracked facial points is put into the classifier. One of six basic emotions is the last output of this algorithm. In the following sections in this chapter, facial point tracking and classifier are explained in details.

3.2 Facial Point Tracking

3.2.1 Eyes and Eyebrows

Points 3, 4, 5 and 6

The real probability distribution is as Eq. (3.1). $\mathcal{N}_k$ is the normal distribution and X_k is the position of facial point k in the current frame. ${}^{-1}X_k$ is the position of facial point k in the previous frame. ${}^{0}X_k$ is the position of facial point k in the first frame which is the frame with the neutral face. $\mathbf{z}$ is gray scale window centered at point X_k and $\mathbf{c}$ is a gray scale template of point k. Δ is the normalized value.

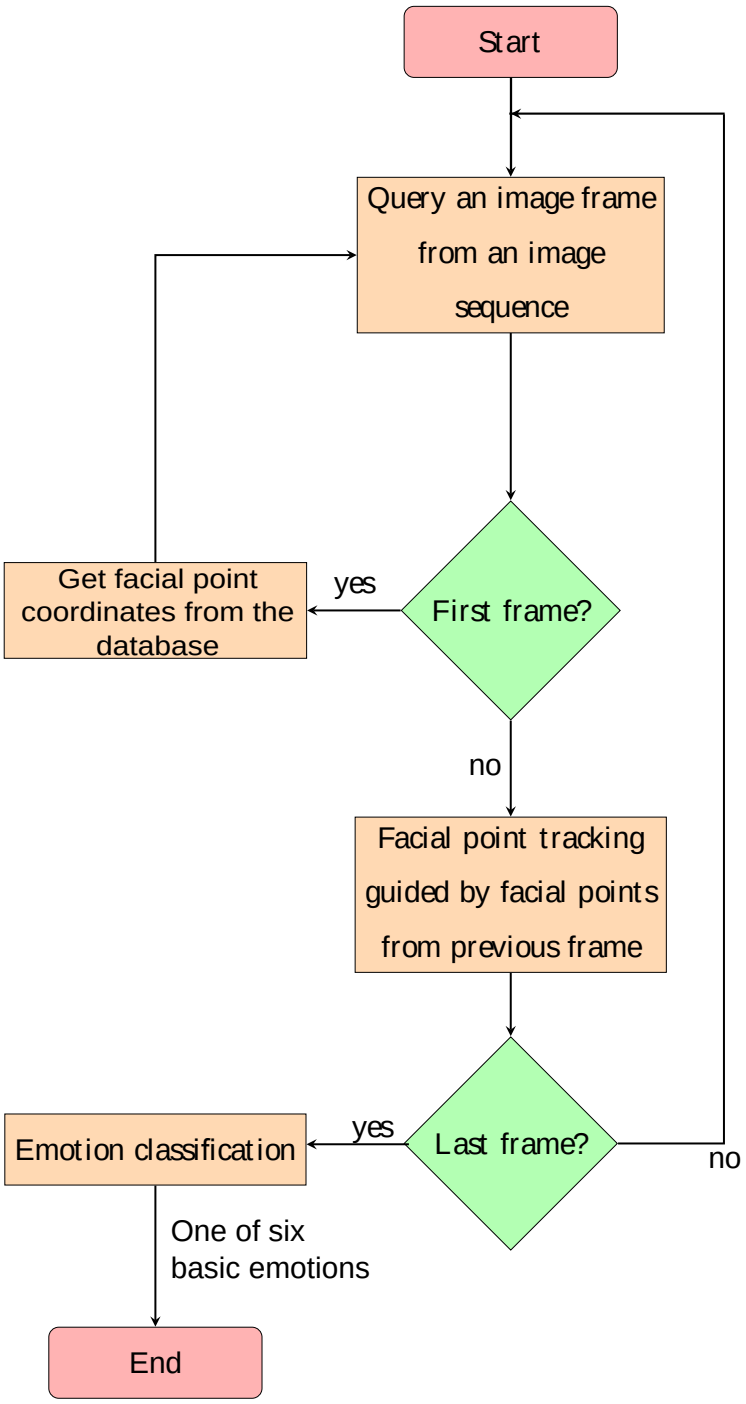

Figure 3.1: Flowchart showing overview of emotion detection algorithm

$$Pr(X = \{X_3, X_4, X_5, X_6\}) = \begin{cases} (\sum_{{}^{-1}X} \mathcal{N}_k({}^{-1}X, X)) p(\mathbf{z}|X_k) & \text{if } X_3,\ X_4,\ X_5,\ X_6 \\ & \text{correspond to equa-} \\ & \text{tions (2.7, 2.8, 2.9,} \\ & \text{2.10, 2.11, 2.12)} \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

where

$$\mathcal{N}_k({}^{-1}X, X) = \prod_{k \in \{3,4,5,6\}} \mathcal{N}(x = X_k, mean = {}^{-1}X_k, scale)$$

$$p(\mathbf{z}|X_k) = \Delta \times e^{\frac{(\rho(\mathbf{c},\mathbf{z})-1)}{0.05}}$$

$$\rho(\mathbf{c}, \mathbf{z}) = \left| \frac{\sum_{r \in R^2}[c(r) - \bar{c}][z(r) - \bar{z}]}{\sqrt{\sum_{r \in R^2}[c(r) - \bar{c}]^2}\sqrt{\sum_{r \in R^2}[z(r) - \bar{z}]^2}} \right|, 0 \leq \rho \leq 1$$

However, the complexity of calculating all the probability values is high because the random variables X_k depend on the random variables in the previous frame ${}^{-1}X_k$. Moreover, the random variables in the current state also depend on other random variables in the current state, e.g. X_3, X_4, X_5, X_6 depend on each other. The N particles can be used to estimate the probability distribution thus Eq. (3.1) can be changed to Eq. (3.2) where i means the i^{th} particle. The explanation of each variable is similar to above. Δ is the value for normalization. Each i^{th} particle is sampled using the product of $\mathcal{N}_k^i$ multiplied by $p(\mathbf{z^i}|\mathbf{X_k^i})$ where k is $3, 4, 5, 6$, respectively. If the 4 positions $3, 4, 5, 6$ of the i^{th} particle correspond to equations (2.7) to (2.12), the weight is set to $p(\mathbf{z^i}|\mathbf{X_k^i})$; otherwise the weight is set to zero. After all particles are calculated, the number of particles with non-zero weight might be less than N so the re-sampling is needed to make all particles have equal weights and the number of particles become N again.

$$Pr(X_A^i, X_B^i, X_C^i, X_D^i) = \begin{cases} \prod_{k \in \{3,4,5,6\}} p(\mathbf{z^i}|\mathbf{X_k^i}) & \text{if } X_3^i, X_4^i, X_5^i, X_6^i \text{ are in} \\ & \text{bounding conditions} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

where

$$X_k^i \text{ is sampled from } \mathcal{N}(mean_k = {}^{-1}X_k^i, scale_k = {}^{-1}X_k^i - {}^{0}X_k^i)$$

$$p(\mathbf{z^i}|\mathbf{X_k^i}) = \Delta \times e^{\frac{(\rho(\mathbf{c},\mathbf{z^i})-1)}{0.05}}$$

$$\rho(\mathbf{c},\mathbf{z^i}) = \left| \frac{\sum_{r\in R^2}[c(r)-\bar{c}][z^i(r)-\bar{z^i}]}{\sqrt{\sum_{r\in R^2}[c(r)-\bar{c}]^2}\sqrt{\sum_{r\in R^2}[z^i(r)-\bar{z^i}]^2}} \right|, 0 \leq \rho \leq 1$$

However, the particle method above has some step that sets some particle weights to zero thus the accuracy of probability may reduce rapidly over time. We propose the new method using Eq. (3.3) to approximate the probabilistic value of Eq. (3.1). By using this method we can avoid to set the weight of some particles to zero by using the conditional probability and sample the i^{th} particle step-by-step from point A to point D where (A, B, C, D) are different among particles. For example, the 1st particle may have (A, B, C, D) as $(3, 4, 5, 6)$ and the 2^{nd} particle may have (A, B, C, D) as $(4, 6, 3, 5)$. We use 96 particles to estimate the probabilistic model. Each particle has eight dimensions that consists of the coordinates x and y of four points 3, 4, 5, and 6. There are four steps (four probabilistic sampling), one step for each facial point, for each particle. Each current particle depends on the particles of the previous frame. Each particle of 100 particles has its own criteria. For example, if the first particle has $(3, 4, 5, 6)$ and the second particle has $(5, 6, 4, 3)$, the first particle will perform the first step to localize the 3^{rd} facial point and the last step for the 6^{th} facial point as well as the second particle will perform the first step to localize the 5^{th} facial point and the last step for the 3^{rd} facial point. The reason that each particle has different criteria is to reduce risk from relying on some criteria too much. For example, if we use only $(3, 4, 5, 6)$ criteria for all particles and the 3^{rd} point position for each particle fails to localize the 3^{rd} facial point position, the rest of the facial point position detection will be affected because the next step accuracy also depend on the previous step. All possible criteria are propagated to the particles equally. A criterion is the member of set $\{(A, B, C, D) \mid A, B, C, D \in \{3, 4, 5, 6\} \wedge A \neq B \neq C \neq D\}$. Each criterion has similar explanation. For example in case of $(3, 4, 5, 6)$, the position of

the 3^{rd} facial point is localized using the initial distribution for facial point 3 and the weight of the particle will be adapted by texture correlation and by possible area of facial points 4, 5, 6 related to founded facial point 3. After the weight is adapted, particles need re-sampling to contain 100 particles again. After the point 3 is processed, the position of the facial point 4 is localized using the conditional probability from point 3 to point 4. Then the position of the facial point 5 is localized using the conditional probability from points 3 and 4 to point 5. Finally, the position of the facial point 6 is localized using the conditional probability from points 3, 4, and 5 to point 6. The geometric conditions for finding these four points can be found in the matrices in Eqs. (3.9, 3.10, 3.11, 3.12, 3.13, 3.14) that are similar to equations (2.7, 2.8, 2.9, 2.10, 2.11, 2.12). we define $z = 0$, $t = \inf$ and $\odot$ is dot product. $\mathcal{T}(LB, UB, mean, scale)$ is the trancated normal distribution where LB is the lower truncation limit, UB is the upper truncation limit, *mean* is the mean of the distribution, and *scale* is the standard deviation of the distribution. t is ∞, z is 0, and ξ is just dummy data. $Ept01$ is the average between x-axis values of facial points 1 and 2. Φ is the golden ratio. $L1$, $U1$, $L2$, and $U2$ are defined in matrix form for other equations reaching its elements easily.

$$\Pr(X_A^i, X_B^i, X_C^i, X_D^i) = \Pr(X_D^i \mid X_C^i, X_B^i, X_A^i) \Pr(X_C^i \mid X_B^i, X_A^i) \Pr(X_B^i \mid X_A^i) \Pr(X_A^i \mid {}^{-1}X_A^i) \Pr({}^{-1}X_A^i) \quad (3.3)$$

$$\Pr({}^{-1}X_A^i, {}^{-1}X_B^i, {}^{-1}X_C^i, {}^{-1}X_D^i) = InitialValue \quad (3.4)$$

$$\Pr(X_A^i \mid {}^{-1}X_A^i) = F_{{}^{-1}X_B^i, {}^{-1}X_C^i, {}^{-1}X_D^i} \times p(\mathbf{z^i}|\mathbf{X^i}_A) \quad (3.5)$$

$$\Pr(X_B \mid X_A, {}^{-1}X_B^i) = F_{X_C^{-1}, X_D^{-1}} \times p(\mathbf{z^i}|\mathbf{X^i}_B) \quad (3.6)$$

$$\Pr(X_C^i \mid X_B^i, X_A^i, {}^{-1}X_C^i) = F_{{}^{-1}X_D^i} \times p(\mathbf{z^i}|\mathbf{X^i}_C) \quad (3.7)$$

$$\Pr(X_D^i \mid X_C^i, X_B^i, X_A^i, {}^{-1}X_D^i) = p(\mathbf{z^i}|\mathbf{X^i}_D) \quad (3.8)$$

where

- X_A^i is sampled from $\mathcal{T}(LB_A^i, UB_A^i, mean_A = {}^{-1}X_A^i, scale_A = {}^{-1}X_A^i - {}^{0}X_A^i)$,
- X_B^i is sampled from $\mathcal{T}(LB_{A2B}^i, UB_{A2B}^i, mean_B = {}^{-1}X_B^i, scale_B = {}^{-1}X_B^i - {}^{0}X_B^i)$,
- X_C^i is sampled from $\mathcal{T}(LB_{AB2C}^i, UB_{AB2C}^i, mean_C = {}^{-1}X_C^i, scale_C = {}^{-1}X_C^i - {}^{0}X_C^i)$,
- X_D^i is sampled from $\mathcal{T}(LB_{ABC2D}^i, UB_{ABC2D}^i, mean_D = {}^{-1}X_D^i, scale_D = {}^{-1}X_D^i - {}^{0}X_D^i)$

$$
\begin{aligned}
Lfn(q,r)^i &= L1_{q,r} \odot X_q^i + L2_{q,r} \\
Ufn(q,r)^i &= U1_{q,r} \odot X_q^i + U2_{q,r} \\
LB_A^i &= {}^{-1}X_A^i + L_A \\
UB_A^i &= {}^{-1}X_A^i + U_A \\
LB_{A2B}^i &= \max(Lfn(A,B)^i, {}^{0}X_B^i + L_B) \\
UB_{A2B}^i &= \min(Ufn(A,B)^i, {}^{0}X_B^i + U_B) \\
LB_{AB2C}^i &= \max(Lfn(A,C)^i, Lfn(B,C)^i, {}^{0}X_C^i + L_C) \\
UB_{AB2C}^i &= \min(Ufn(A,C)^i, Lfn(B,C)^i, {}^{0}X_C^i + U_C) \\
LB_{ABC2D}^i &= \max(Lfn(A,D)^i, Lfn(B,D)^i, Lfn(C,D)^i, {}^{0}X_D^i + L_D) \\
UB_{ABC2D}^i &= \min(Ufn(A,D)^i, Ufn(B,D)^i, Ufn(C,D)^i, {}^{0}X_D^i + U_D)
\end{aligned}
$$

$$
\begin{aligned}
F_{{}^{-1}X_B^i, {}^{-1}X_C^i, {}^{-1}X_D^i} = &\prod_{k \in \{B,C,D\}} \Pr(LB_{A2k}^i \le {}^{-1}X_k^i \le UB_{A2k}^i) \\
&\Pr({}^{0}X_A^i + L_A \le X_A^i \le {}^{0}X_A^i + U_A) \\
F_{{}^{-1}X_C^i, {}^{-1}X_D^i} = &\prod_{k \in \{C,D\}} \Pr(LB_{B2k}^i \le {}^{-1}X_k^i \le UB_{B2k}^i) \\
F_{{}^{-1}X_D^i} = &\prod_{k \in \{D\}} \Pr(LB_{C2k}^i \le {}^{-1}X_k^i \le UB_{C2k}^i)
\end{aligned}
$$

$$L1 = \begin{array}{c} \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccc} 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} \xi & (-1,z) & (1,1) & (z,z) \\ (-1,z) & \xi & (z,z) & (1,1) \\ (1,z) & (z,z) & \xi & (-1,z) \\ (z,z) & (1,1) & (-1,z) & \xi \end{bmatrix} \end{array} \tag{3.9}$$

$$U1 = \begin{array}{c} \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccc} 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} \xi & (-1,t) & (1,t) & (t,t) \\ (-1,t) & \xi & (t,t) & (1,t) \\ (1,1) & (t,t) & \xi & (-1,t) \\ (t,t) & (1,t) & (-1,t) & \xi \end{bmatrix} \end{array} \tag{3.10}$$

$$L2 = \begin{array}{c} \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cc} 3 & 4 \end{array} \\ \begin{bmatrix} \xi & [2(-0.5EMW + Ept01), 0] \\ [2(-0.5EMW + Ept01), 0] & \xi \\ [\frac{2EW}{3}, 0] & [0,0] \\ [0,0] & [\frac{-4EW}{3}, \frac{-EH}{3}]] \end{array} \\ \begin{array}{cc} 5 & 6 \end{array} \\ \begin{bmatrix} [\frac{-4EW}{3}, \frac{-EH}{3}] & [0,0] \\ [0,0] & [\frac{2EW}{3}, 0] \\ \xi & [2 * (-0.5EMW + Ept01), 0] \\ [2(-0.5EMW + Ept01), 0] & \xi \end{bmatrix} \end{array} \tag{3.11}$$

$$U2 = \begin{array}{c} \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cc} 3 & 4 \end{array} \\ \begin{bmatrix} \xi & [2(0.5EMW + Ept01), 0] \\ [2(0.5EMW + Ept01), 0] & \xi \\ [\frac{4EW}{3}, \frac{EH}{3}] & [0, 0] \\ [0, 0] & [\frac{-2EW}{3}, 0] \end{bmatrix} \end{array}$$

$$\begin{array}{c} \begin{array}{cc} 5 & 6 \end{array} \\ \begin{bmatrix} [\frac{-2EW}{3}, 0] & [0, 0] \\ [0, 0] & [\frac{4EW}{3}, \frac{EH}{3}] \\ \xi & [2(0.5EMW + Ept01), 0] \\ [2(0.5EMW + Ept01), 0] & \xi \end{bmatrix} \end{array} \tag{3.12}$$

$$L = \begin{array}{c} \begin{array}{cccc} 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} [-(\Phi - 1)ew_m, -\frac{ew}{1.5}] & [-ew_m, -\frac{ew}{1.5}] & [-ew_m, -\frac{ew}{1.5}] & [-(\Phi - 1)ew_m, -\frac{ew}{1.5}] \end{bmatrix} \end{array} \tag{3.13}$$

$$U = \begin{array}{c} \begin{array}{cccc} 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} [ew_m, \frac{ew}{1.5}] & [(\Phi - 1)ew_m, \frac{ew}{1.5}] & [(\Phi - 1)ew_m, \frac{ew}{1.5}] & [ew_m, \frac{ew}{1.5}] \end{bmatrix} \end{array} \tag{3.14}$$

and

$$ew = \frac{EW}{2.5}$$

$$ew_m = 1.25ew$$

After all particles are processed, the mean of each particle is assigned to be an accurate position of each facial point as shown in Eq. (3.15). $x(s)$ means the assigned positions of point s.

$$x(s) = \frac{1}{N_s}\Sigma_{i=1}^{N_s} X_s^i \text{ where } s \in 3, 4, 5, 6 \tag{3.15}$$

Points 7 and 8

After points 3,4,5 and 6 are assigned, point 7 is localized using Eq. (3.16)

$$\Pr(X_7^i) = p(\mathbf{z^i}|\mathbf{X_7^i}) \tag{3.16}$$

X_7^i is sampled from $\mathcal{T}(LB_7^i, UB_7^i, mean_7 = {}^{-1}X_7^i, scale_7 = {}^{-1}X_7^i - {}^{0}X_7^i)$

where

$$LB_7^i = [W\Big(\frac{x(3)+x(5)}{2}\Big) - \frac{1}{4} \times EW, Height\Big(\frac{x(3)+x(5)}{2}\Big)]$$

$$UB_7^i = [W\Big(\frac{x(3)+x(5)}{2}\Big) + \frac{1}{4} \times EW, Height\Big(\frac{x(3)+x(5)}{2}\Big) + EH]$$

Equation of point 8 is defined similar to Eq. (3.16).

Points 11, 12, 13, and 14

After points 7 and 8 are assigned, points 11,12,13 and 14 are localized using Eq, (3.17) to Eq. (3.20) and geometric features (3.2.1, 3.2.1, 3.2.1, 3.2.1, 3.2.1). A criteria for each particle is the member of set $\{(E,F) \mid E,F \in \{11,13\} \wedge E \neq F\}$.

$$\Pr(X_E^i, X_F^i) = \Pr(X_F^i \mid X_E^i, {}^{-1}X_F^i)\Pr(X_E^i \mid {}^{-1}X_E^i)\Pr({}^{-1}X_E^i, {}^{-1}X_F^i) \tag{3.17}$$

where

$$\Pr({}^{-1}X_E^i, {}^{-1}X_F^i) = InitialValue \tag{3.18}$$

$$\Pr(X_E^i \mid {}^{-1}X_E^i) = F_{{}^{-1}X_F^i} \times p(\mathbf{z^i}|\mathbf{X_E^i}) \tag{3.19}$$

$$\Pr(X_F \mid X_E, {}^{-1}X_F^i) = p(\mathbf{z^i}|\mathbf{X_F^i}) \tag{3.20}$$

X_E^i is sampled from $\mathcal{T}(LB_E^i, UB_E^i, mean_E = {}^{-1}X_E^i, scale_E = {}^{-1}X_E^i - {}^{0}X_E^i)$
X_F^i is sampled from $\mathcal{T}LB_{E2F}^i, UB_{E2F}^i, mean_F = {}^{-1}X_F^i, scale_F = {}^{-1}X_F^i - {}^{0}X_F^i)$

$$
\begin{aligned}
LB_E^i &= {}^{-1}X_E^i + L_E \\
UB_E^i &= {}^{-1}X_E^i + U_E \\
LB_{E2F}^i &= \max(Lfn(E,F)^i, {}^{-1}X_F^i + L_F) \\
UB_{E2F}^i &= \min(Ufn(E,F)^i, {}^{-1}X_F^i + U_F) \\
F_{{}^{-1}X_F^i} &= \Pr(LB_{E2F}^i \leq {}^{-1}X_F^i \leq UB_{E2F}^i) \Pr({}^{0}X_E^i + L_E \leq X_E^i \leq {}^{0}X_E^i + U_E)
\end{aligned}
$$

$$
L1_{eb} = \begin{matrix} \\ 3 \\ 4 \end{matrix} \begin{matrix} \begin{matrix} 11 & 13 \end{matrix} \\ \begin{bmatrix} \xi & (1,z) \\ (1,z) & \xi \end{bmatrix} \end{matrix}
$$

$$
U1_{eb} = \begin{matrix} \\ 3 \\ 4 \end{matrix} \begin{matrix} \begin{matrix} 11 & 13 \end{matrix} \\ \begin{bmatrix} \xi & (1,t) \\ (1,t) & \xi \end{bmatrix} \end{matrix}
$$

$$
L2_{eb} = \begin{matrix} \\ 3 \\ 4 \end{matrix} \begin{matrix} \begin{matrix} 11 & 13 \end{matrix} \\ \begin{bmatrix} \xi & [-\frac{4}{3}EBW, 0] \\ [\frac{2}{3}EBW, 0] & \xi \end{bmatrix} \end{matrix}
$$

$$
U2_{eb} = \begin{matrix} \\ 3 \\ 4 \end{matrix} \begin{matrix} \begin{matrix} 11 & 13 \end{matrix} \\ \begin{bmatrix} \xi & [-\frac{2}{3}EBW, 0] \\ [\frac{4}{3}EBW, 0] & \xi \end{bmatrix} \end{matrix}
$$

$$
L_{eb} = \begin{matrix} \begin{matrix} 11 & 13 \end{matrix} \\ \begin{bmatrix} [-20,-20] & [-20,-20] \end{bmatrix} \end{matrix}
$$

$$
U_{eb} = \begin{matrix} \begin{matrix} 11 & 13 \end{matrix} \\ \begin{bmatrix} [20,20] & [20,20] \end{bmatrix} \end{matrix}
$$

Equations of points 12 and 14 are defined similar to Eq. (3.21).

Points 9 and 10

After points 11,12,13 and 14 are assigned, point 9 is localized using Eq. (3.21)

$$\Pr(X_9^i) = p(\mathbf{z^i}|\mathbf{X_9^i}) \tag{3.21}$$

X_9^i is sampled from $\mathcal{T}(LB_9^i, UB_9^i, mean_9 = {}^{-1}X_9^i, scale_9 = {}^{-1}X_9^i - {}^{0}X_9^i)$

where

$$LB_9^i = [W\Big(\frac{x(3)+x(5)}{2}\Big) - \frac{1}{4} \times EW, Height(x(11))]$$
$$UB_9^i = [W\Big(\frac{x(3)+x(5)}{2}\Big) + \frac{1}{4} \times EW, Height(x(7))]$$

Equation of point 10 is defined similar to Eq. (3.21).

3.2.2 Mouth

Point 15

After points 9 and 10 are assigned, point 9 is localized using Eq. (3.22)

$$\Pr(X_{15}^i) = p(\mathbf{z^i}|\mathbf{X_{15}^i}) \tag{3.22}$$

X_{15}^i is sampled from $\mathcal{T}(LB_{15}^i, UB_{15}^i, mean_{15} = {}^{-1}X_{15}^i, scale_{15} = {}^{-1}X_{15}^i - {}^{0}X_{15}^i)$

where

$$LB_{15}^i = [Width(x(1)), Height\Big(\frac{x(3)+x(5)}{2}\Big)]$$
$$UB_{15}^i = [Height\Big(\frac{x(1)+x(2)}{2}\Big) + \frac{1}{2} \times HMN, \inf]$$

Points 17 and 18

After point 15 is assigned, points 17 and 18 are localized using eq. (3.23).

$$\Pr(X_{17}^i) = p(\mathbf{z^i}|\mathbf{X_{17}^i}) \tag{3.23}$$

X_{17}^i is sampled from $\mathcal{T}(LB_{17}^i, UB_{17}^i, mean_{17} = {}^{-1}X_{17}^i, scale_{17} = {}^{-1}X_{17}^i - {}^{0}X_{17}^i)$

where

$$LB_{17}^i = [Width\Big(x(15)\Big) - 2 \times MW, Height(x(15))]$$

$$UB_{17}^{i} = [Width\Big(x(15)\Big) - \frac{1}{2} \times MW, \inf]$$

Equations of point 18 is defined similar to Eq.(3.23).

Point 16

After point 15 is assigned, point 16 is localized using Eq. (3.24).

$$\Pr(X_{16}^{i}) = p(\mathbf{z^i}|\mathbf{X_{16}^i}) \tag{3.24}$$

X_{16}^{i} is sampled from $\mathcal{T}(LB_{16}^{i}, UB_{16}^{i}, mean_{16} = {}^{-1}X_{16}^{i}, scale_{16} = {}^{-1}X_{16}^{i} - {}^{0}X_{16}^{i})$

where

$$LB_{16}^{i} = \Big[Width(x(1)), Height\Big(\frac{x(17)+x(18)}{2}\Big)\Big]$$
$$UB_{16}^{i} = [Width(x(2)), \inf]$$

$$\Pr(X_{18}^{i}) = p(\mathbf{z^i}|\mathbf{X_{18}^i}) \tag{3.25}$$

where

X_{18}^{i} is sampled from $\mathcal{T}(LB_{18}^{i}, UB_{18}^{i}, mean_{18} = {}^{-1}X_{18}^{i}, scale_{18} = {}^{-1}X_{18}^{i} - {}^{0}X_{18}^{i})$

$$LB_{18}^{i} = \Big[Width\Big(x(15)\Big) + \frac{1}{2} \times MW, Height\Big(\frac{x(3)+x(5)}{2}\Big), Height(x(15))\Big]$$
$$UB_{18}^{i} = \Big[Width\Big(x(15)\Big) + 2 \times MW, \inf\Big]$$

3.2.3 Blob and Binary Information

For some facial points, we do not use only the spatial and texture model as described above in sections 3.2.1 and 3.2.2 to localize the facial points but also use the blob region described in section 2.2.5 to reduce the possible region. As seen in figure 3.2, we bound the area to find blobs for eyebrows. The green one is the example of founded blob that is the largest blob in this rectangular area. This founded blob can help to reduce the area of finding points 18 and 21.

Considering only the region containing mouth without eyes and a nose, *blob*s are extracted as seen in Figure 2.9 to help localization of points 20 and 21. Let

Figure 3.2: Example of founded blob shown in green color for eyebrow localization

$blob \in \overline{\beta}$ where $\overline{\beta}$ is the set of founded *blob*s. We use the condition from Eq. (3.26) to eliminate irrelevant *blob*s. h_0, h_1, v_0, and v_1 can be defined as seen in Figure 3.3. $contour(blob)$ is composed of the edge points of a *blob*. Let $\beta \subseteq \overline{\beta}$ where β is a set of candidate *blob*s. Let $(h, v) \in contour(blob)$, where h and v is the horrizontal and vertical coordinates of the edge points of the *blob*.

$$\beta = \Bigg\{ blob \mid (h,v) \in contour(blob), h_0 \leq h \leq h_0 + \frac{h_1 - h_0}{3},\ v_0 \leq v \leq v_1, \\ h_0 + \frac{h_1 - h_0}{3} < h < h_0 + 2\left(\frac{h_1 - h_0}{3}\right),\ v_0 \leq v \leq v_1, \\ h_0 + 2\left(\frac{h_1 - h_0}{3}\right) \leq h \leq h_1,\ v_0 \leq v \leq v_1 \Bigg\} \quad (3.26)$$

3.3 Strong Points of Using Probabilistic Model for Facial Point Tracking

Low complexity Complexity of probabilistic model used in this research is $O(n)$ where n is the number of particles used for probabilistic value estimation. Because of low complexity, cost of computing is effect only a bit while the dimensions of the model increase.

No training process Some methods for facial point tracking need information from many image sequences and many frames to train the tracking model.

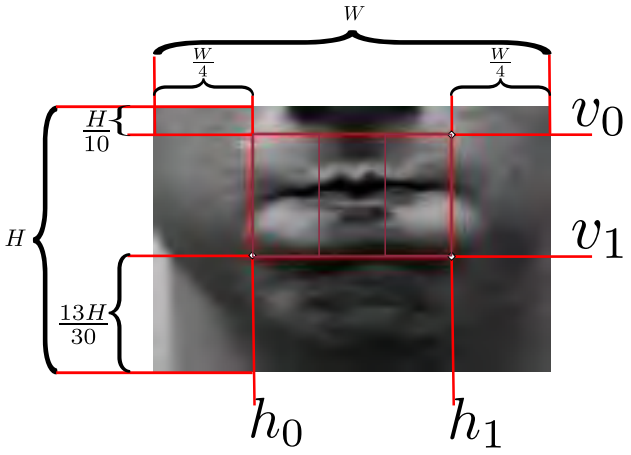

Figure 3.3: Definition of h_0, h_1, v_0, and v_1 that are used in finding candidate *blob*s process for mouth

The benefits of those methods are that they have high speed and high accuracy for facial point tracking. However, the accuracy for trained model usually has high accuracy for only the training data set and has low accuracy for the unseen or new data set. In contrast, our model do not need the training process. Our model can solve this problem by using only the texture from the neutral face of an image sequence to form the model so that it is easier to adapt our method to a new image sequence.

3.4 Emotion Classification

After all facial points as seen in Figure 2.7 are localized we will choose only the first and the last frame of each image sequence to create features.

3.4.1 Features

The first frame contains a neutral face. The last frame contains a face with peak emotion. The distances between points are used to create features so a number of features for 18 facial points of the first frame is $\binom{18}{2}$ and for the last frame is $\binom{18}{2}$ features as well. After features from the first and the last frame are extracted, they are subtracted with each other as Eq. (3.27) for all i to produce the new

feature.

$$fea_i = \|fea_i^0 - fea_i^{last}\| \tag{3.27}$$

where

$$i \in \{1, 2, ..., \binom{18}{2}\}$$
$$fea_i^{fr} \text{ is the } i^{th} \text{ feature of frame } fr$$

To avoid redundant of features, some distances that have the same meaning such as height distances of eyes from point 7 to 9 and distance from point 8 to 10 will be merged by averaging these two distances together, thus the number of features will be reduced from $\binom{18}{2}$ to M features. The reasons for choosing only differences of facial distances from first and last frame of the image sequence to produce the features is that

- Last frame contains the face with a peak of emotion. The unique facial distances are seen from the face with a peak of each emotion easier than the faces from other frames in the image sequence,
- Using many frames can give more information to the classifier but it can produce more dimensions of features that are related to feature redundancy, classification overfitting, and speed of computing issues.

3.4.2 Feature Selection and Classification

After the number of features are reduced to M, we use the classifier and feature selection from orange library. Orange is an open source library for data mining through visual programming or Python scripting. We choose SVM as the classifier and we use the relief method for feature selection to reduce features from M features to 40 features to avoid over fitting. We use 10 folds cross validation to test the performance of our model. The parameters of the SVM are set as seen in Table 3.1. The parameters of SVM are adjusted automatically by Orange library for increase of classification rate.

Parameters	Methods
kernel	RBF
mullticlass stratigie	one vs one
other parameters	automatically adjusted by orange library

Table 3.1: SVM parameters used in our research

CHAPTER IV

Results and Discussion

4.1 Facial Point Tracking

We show the results of some image sequences for the first, the middle and the last frames in each emotion. We show facial tracking with high accuracy and some results with low accuracy as well.

4.1.1 Accurate Tracking

The tracking results of the sequences labeled as one of the basic emotions are shown in Figures 4.1, 4.2, 4.3, 4.4, 4.5, and 4.6. The results of three image sequences are shown for each emotion. The first, the second, and the third columns in the Figures are the first, the middle, and last frames of the image sequences, respectively. Tracking results of 18 facial points around eyebrows, eyes, nose, and mouth are shown in each figure.

4.1.2 Inaccurate Tracking

Some of image sequences have inaccurate tracking results as shown in Figures 4.7 and 4.8. The first, the second, and the third columns in the Figures are the first, the middle, and last frames of the image sequences, respectively. The reasons of inaccurate tracking are that

- Using particles for probabilistic value approximation cause accuracy to drop but it reduces complexity of calculation.
- Face deformation of emotion expression is not rigid so that using the static templates form the neutral face causes accuracy to drop when deformation of

Figure 4.1: This figure shows tracking results of three image sequences that are labeled as anger. Each row is for each image sequence. The first, the second, and the third columns are the first, the middle, and last frames of the image sequences, respectively.

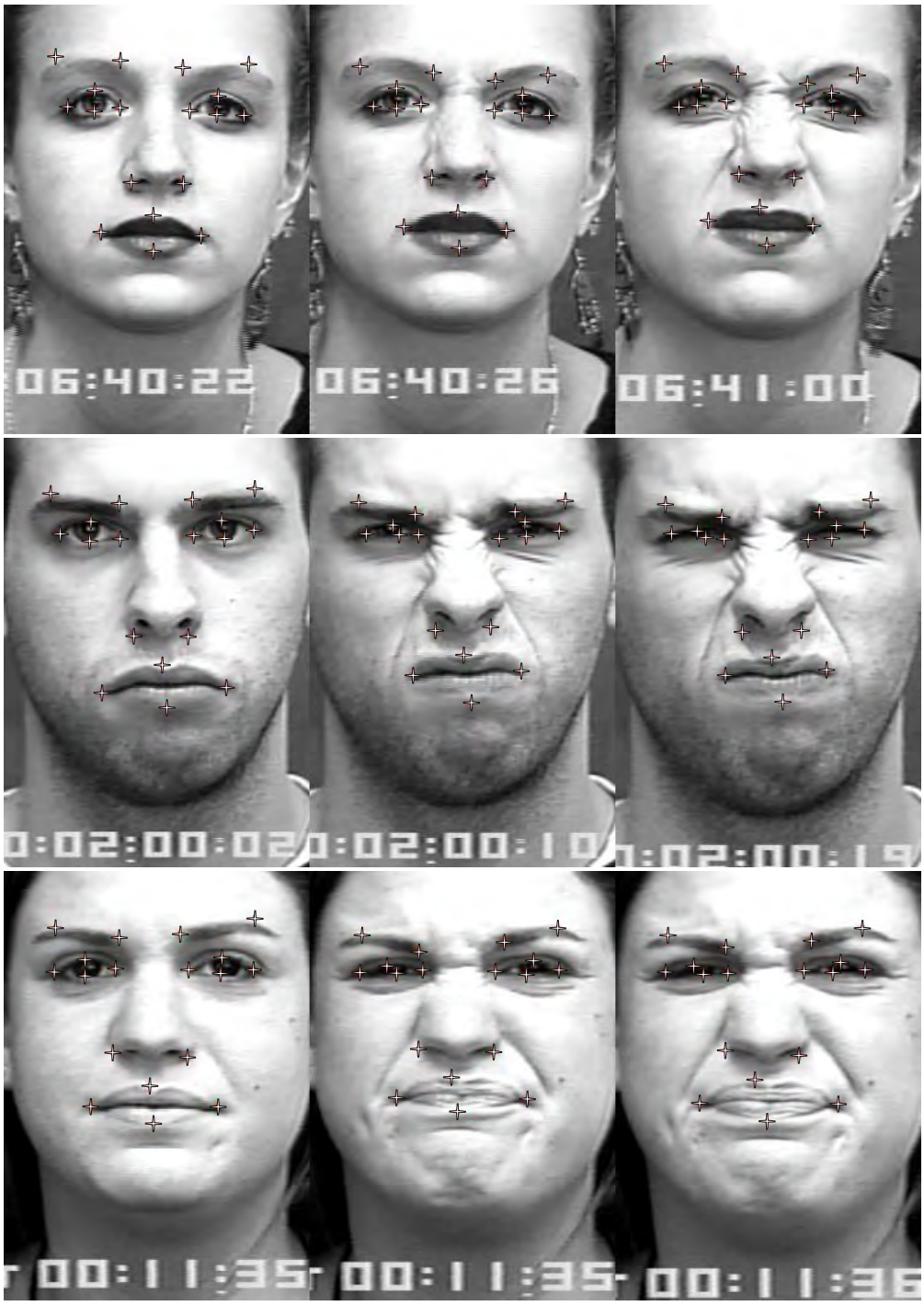

Figure 4.2: This figure shows tracking results of three image sequences that are labeled as disgust. Each row is for each image sequence. The first, the second, and the third columns are the first, the middle, and last frames of the image sequences, respectively.

Figure 4.3: This figure shows tracking results of three image sequences that are labeled as fear. Each row is for each image sequence. The first, the second, and the third columns are the first, the middle, and last frames of the image sequences, respectively.

Figure 4.4: This figure shows tracking results of three image sequences that are labeled as happiness. Each row is for each image sequence. The first, the second, and the third columns are the first, the middle, and last frames of the image sequences, respectively.

Figure 4.5: This figure shows tracking results of three image sequences that are labeled as sadness. Each row is for each image sequence. The first, the second, and the third columns are the first, the middle, and last frames of the image sequences, respectively.

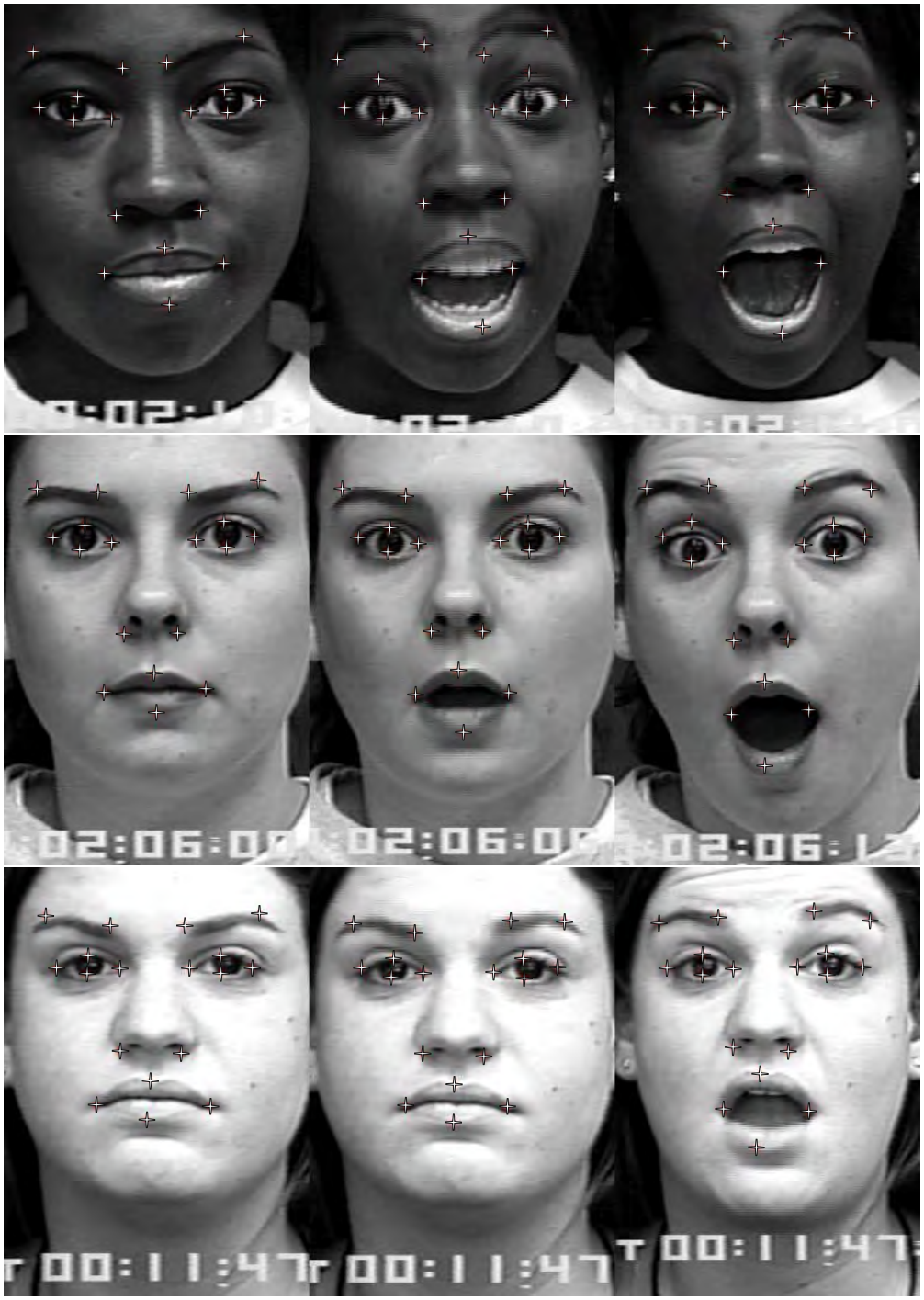

Figure 4.6: This figure shows tracking results of three image sequences that are labeled as surprise. Each row is for each image sequence. The first, the second, and the third columns are the first, the middle, and last frames of the image sequences, respectively.

Real/Predict	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	17	6	0	9	3	1
Disgust	6	43	2	5	1	0
Fear	3	2	7	3	5	3
Happiness	5	7	0	49	2	1
Sadness	1	1	2	3	10	5
Surprise	1	1	0	0	2	71

Table 4.1: Performance of classification

Real/Predict	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	47.22	16.67	0	25	8.33	2.78
Disgust	10.53	75.44	3.51	8.77	1.75	0
Fear	13.04	8.7	30.43	13.04	21.74	13.04
Happiness	7.81	10.94	0	76.56	3.13	1.56
Sadness	4.55	4.55	9.09	13.64	45.45	22.73
Surprise	1.33	1.33	0	0	2.67	94.67

Table 4.2: Performance of classification calculated as percentage of successful prediction

the emotional frame make huge difference from the neutral frame. However, using simple templates like this causes has a complexity of computing.

4.2 Classification Rate

Table 4.1 shows the success rate of emotion classification. The row labels are the real labels of emotions from video sequences and the column labels are the predicted emotions. Table 4.2 shows the Table 4.1 in the form of percentage. In accuracy of classification can come from the inaccurate tracking results and the number of samples. As seen in the result Table 4.1, the results of emotion with low number of samples have low accuracy.

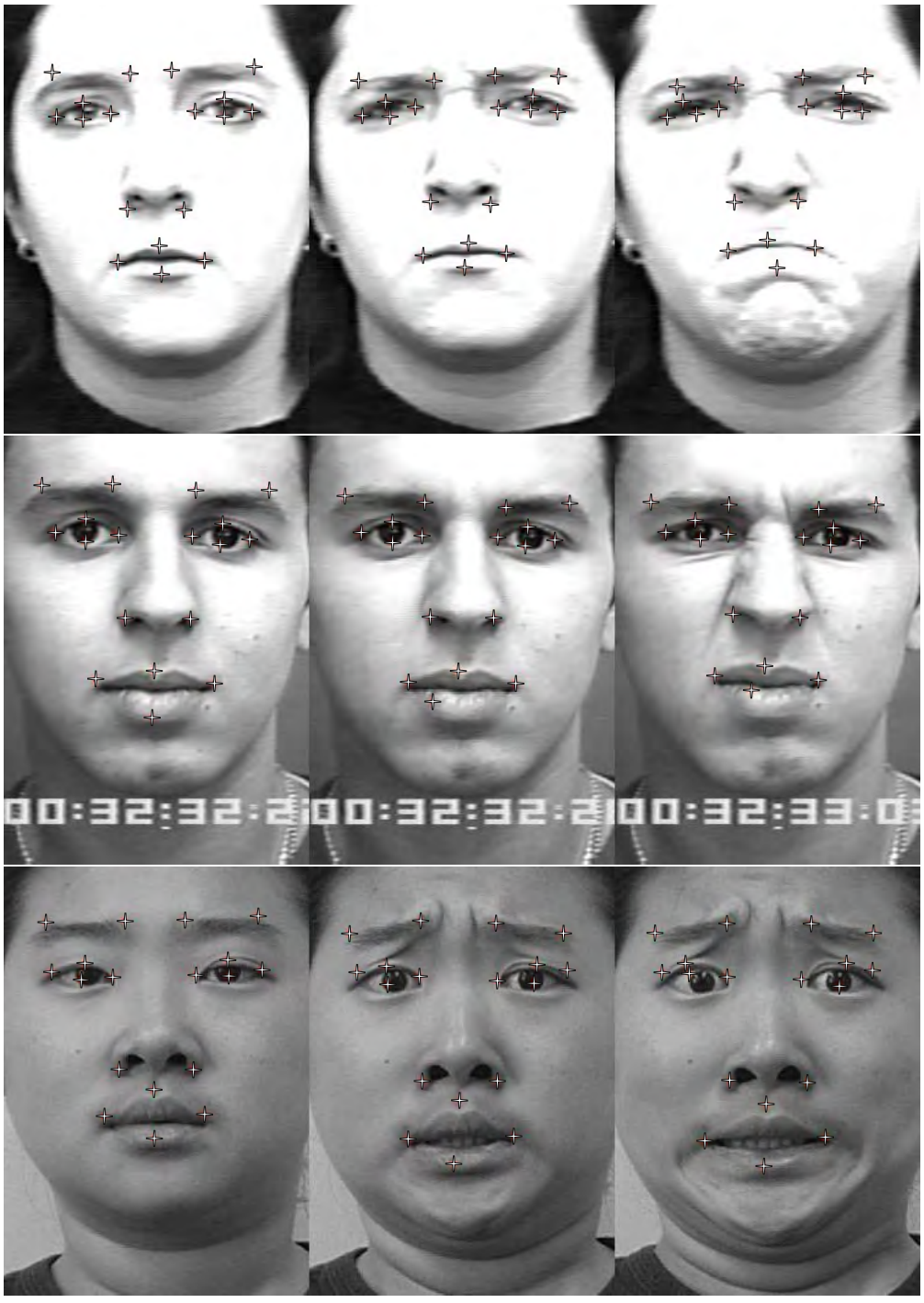

Figure 4.7: This figure shows inaccurate tracking results of some image sequences. Each row is for each image sequence. The first, the second, and the third columns are the first, the middle, and last frames of the image sequences, respectively.

Figure 4.8: This figure shows inaccurate tracking results of some image sequences. Each row is for each image sequence. The first, the second, and the third columns are the first, the middle, and last frames of the image sequences, respectively.

4.3 Discussion

In practical applications, speed and accuracy are important keys. Our method is based on probabilistic estimation so the complexity is low. Moreover, the texture information from the first frame plays an important role for accuracy in this work for facial point tracking. However, this kind of texture is not presented in unseen faces. This is the big limitation that needs to be improved so that emotion detection can be easier adapted to a real world application. We suggest that the future works should focus on tracking algorithm that can localize facial points from any single frame in an image sequence without using the texture information from first frame or the spacial information from the previous frame. Another point is that using the low-level programming language is suggested to make the algorithms run faster.

REFERENCES

[1] Vukadinovic D. and Pantic M.: Fully automatic facial feature point detection using gabor feature based boosted classifiers. *2005 IEEE International Conference on Systems, Man and Cybernetics* 2 (2005): 1692–1698.

[2] Valstar M., Martinez B., Binefa X., and Pantic M.: Facial point detection using boosted regression and graph models. *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2010): 2729–2736.

[3] Fazli S., Afrouzian R., and Seyedarabi H.: Fiducial facial points tracking using particle filter and geometric features. *2010 International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* (2010): 396–400.

[4] Kudiri K., Said A., and Nayan M.: Emotion detection using sub-image based features through human facial expressions. *2012 International Conference on Computer Information Science (ICCIS)* 1 (2012): 332–335.

[5] Saeed A., Al-Hamadi A., Niese R., and Elzobi M.: Effective geometric features for human emotion recognition. *2012 IEEE 11th International Conference on Signal Processing (ICSP)* 1 (2012): 623–627.

[6] Bouvier C., Coulon P.Y., and Maldague X.: Unsupervised lips segmentation based on roi optimisation and parametric model. *IEEE International Conference on Image Processing, 2007. ICIP 2007.* 4 (Sept 2007): IV – 301–IV – 304.

[7] Seyedarabi H., Lee W., and Aghagolzadeh A.: Automatic lip tracking and action units classification using two-step active contours and probabilistic neural networks. *Canadian Conference on Electrical and Computer Engineering, 2006. CCECE '06.* (May 2006): 2021–2024.

[8] Jirajaras T. and Lipikorn R.: Mouth features localization using probability model from blob finding and morphology gradient method. *19th Annual Meeting in Mathematics (AMM2014)* (2014): 153–162.

[9] Lucey P., Cohn J., Kanade T., Saragih J., Ambadar Z., and Matthews I.: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2010): 94–101.

BIOGRAPHY

Name	Thada Jirajaras
Date of Birth	27 August 1989
Place of Birth	Bangkok, Thailand
Education	B.Eng. (Biomedical Engineering), (Second Class Honors), Mahidol University, 2011
Publication	Jirajaras T. and Lipikorn R.: *Mouth features localization using probability model from blob finding and morphology gradient method.*, 19th Annual Meeting in Mathematics, AMM2014, Thailand, pp153–162.