Contents

Chapter 2

Background and Literature Review

The importance of facial expression in social interaction and social intelligence is widely recognized. Facial expression analysis has been an active research topic since 19th century. The first automatic facial expression recognition system was introduced in 1978 by Suwa et al. [83]. This system attempts to analyze facial expressions by tracking the motion of 20 identified spots on an image sequence. Since then, a lot of work has been done in this domain. Various computer systems have been made to help us understand and use this natural form of human communication.

This chapter reviews the state of the art of what has been done in processing and understanding facial expression. When building an FER system, these main issues must be considered: face detection and alignment, image normalization, feature extraction, and classification. Most of the current work in FER is based on methods that implement these steps sequentially and independently. Before exploring what has been done in literature for implementing these steps, we will briefly describe the problem space for facial expression analysis.

2.1 Problem Space for Facial Expression Analysis

2.1.1 Level of Description

In general there are two types of method to describe facial expression.

Facial Action Coding System

The facial action coding system [24] is a human-observer-based system widely used in psychology to describe subtle changes in facial features. FACS consists of 44 action units which are related to contraction of a specific set of facial muscles (Fig.2.1). Some of the action units are shown in Fig.2.2. Conventional, FACS code is manually labeled by trained observers while viewing videotaped facial behavior in slow motion. In recent years, some attempts have been made to do this automatically [69]. The advantage of FACS is its ability to capture the subtlety of facial expression, however FACS itself is purely descriptive and includes no inferential labels. That means in order to get the emotion estimation, the FACS code needs to be converted into the Emotional Facial Action System (EMFACS [28]) or similar systems.

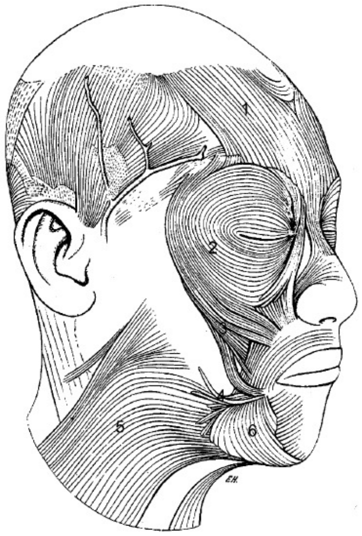

Figure 2.1: Muscles of facial expression. 1, frontalis; 2, orbicularis oculi; 3, zygomaticus major; 4, risorius; 5, platysma; 6, depressor anguli oris [33]

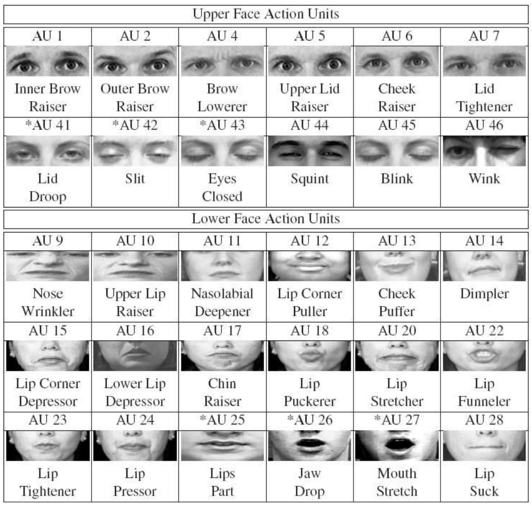

Figure 2.2: FACS action units [35]

Prototypic Emotional Expressions

Instead of describing the detailed facial features, most FER systems attempt to recognize a small set of prototypic emotional expressions. The most widely-used set is perhaps human universal facial expressions of emotion which consists of six basic expression categories that have been shown to be recognizable across cultures 2.3 .

These expressions, or facial configurations have been recognized in people from widely divergent cultural and social backgrounds, and they have been observed even in the faces of individuals born deaf and blind.

These 6 basic emotions, *i.e.*, disgust, fear, joy, surprise, sadness and anger plus "neutral" which means no facial expression are considered in this work. Given a facial image, our system either works as a conventional classifier to determine the most likely emotion or estimates the weights (or possibility) of each emotion as a fuzzy classifier does.

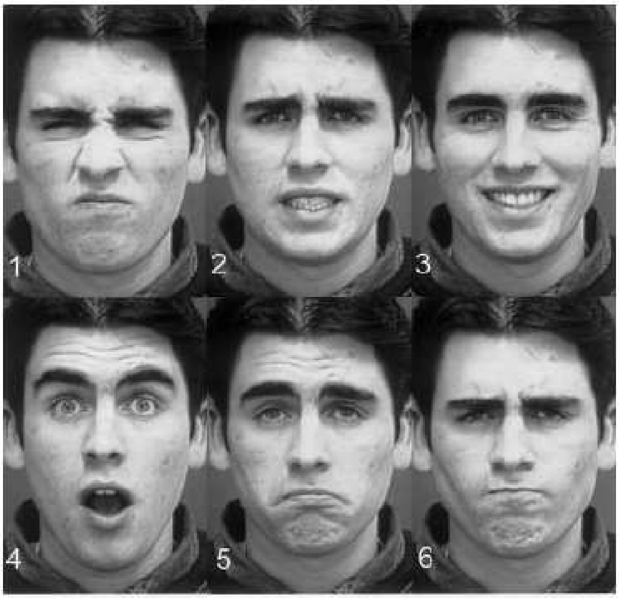

Figure 2.3: Basic facial expression phenotypes. 1, disgust; 2, fear; 3, joy; 4, surprise; 5, sadness; 6, anger

2.2 System Structure

FER can be considered as a special face recognition system or a module of a face recognition system. So it should be instructive to look at the general architecture of a face recognition system. Normally, it consists of four components as depicted in 2.4

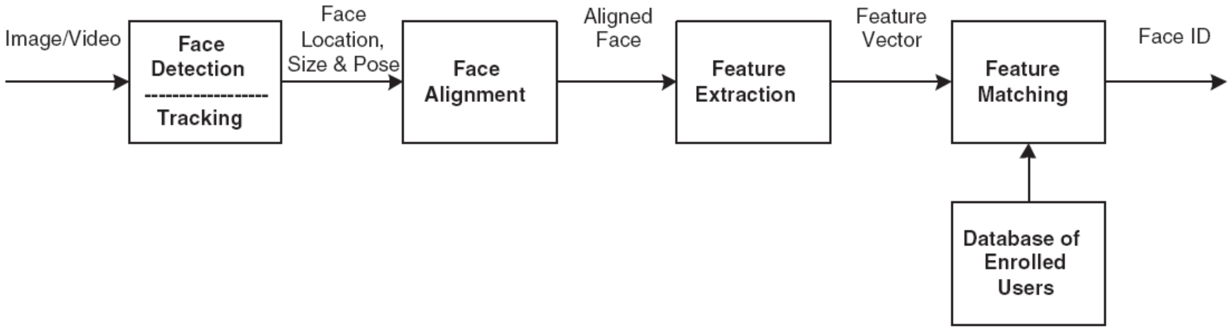

Figure 2.4: Face recognition processing flow

Face detection finds the face areas in the input image. If the input is a video, to be more efficient and also to achieve better robustness, face detection is only performed on key frames and a tracking algorithm is applied on interval frames. Face alignment is very similar to detection, but it is aimed at achieving a more accurate localization. In this step, a set of facial landmarks (facial components), such as eyes, brows and nose, or the facial contour are located; based on that, the face image is rotated, chopped, resized and even warped, this is called geometrical normalization. Usually the face is further normalized with respect to photometrical properties such as illumination and gray scale.

Feature extraction is performed on a normalized face to provide effective information that should be useful for recognizing and classifying labels in which there is interest, such as identity, gender, or expression. The extracted feature vector is sent to a classifier and compared with the training data to produce a recognition output.

2.3 Face Detection

Face detection is the first step in face recognition. It has a major influence on the performance of the entire system. Several cues can be used for face detection, for example, skin color, motion (for videos), facial/head shape, and facial appearance. Most successful face detection algorithms are based on only appearance. This may be because

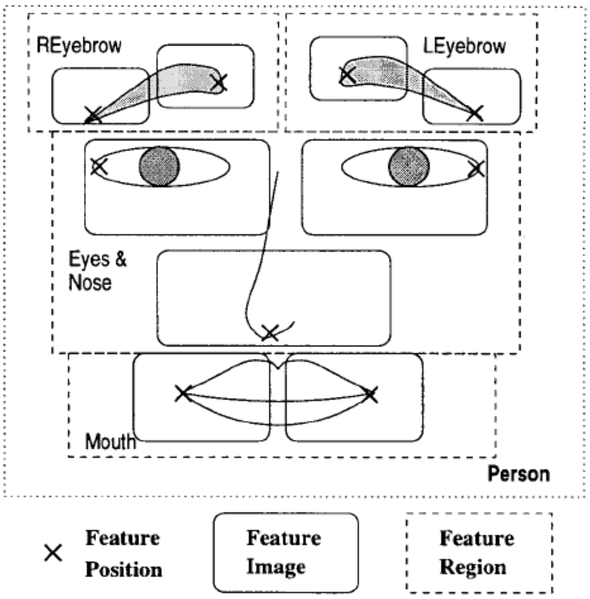

Figure 2.5: Figure 2.12: Scheme of the Facial Features and Regions [15]

2.3.1 Support Vector Machine

Support Vector Machine attempts to construct a linear classifier which maximizes the margin between two classes, so its also known as Optimal Margin Classifier [9]. Fig.2.13 gives an SVM classifier where $\frac{1}{|w|}$ gives the margin and samples along the hyper-planes are called the support vectors.

It has been proven that SVM minimizes the Structural Risk Function which is considered as a better error estimation than the normallyused Empirical Risk Function in terms of generalization capacity.

We consider data points of the form: $\{(x_1, y_1), ..., (x_n, y_n)\}$ where y_i is either 1 or -1, a label denoting the class to which the point x_i belongs. The basic version of SVM can be written as

$$\begin{array}{rl} argmax & \frac{1}{||w||} \\ s.t. & y_i(x^T w - b_0) \geq 1, \ \ \forall i \end{array}$$

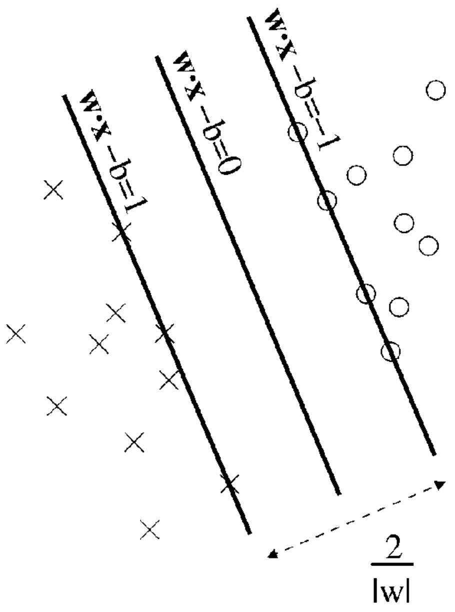

Figure 2.6: Figure 2.13: Maximum-margin hyper-planes for a SVM trained with samples from two classes [99]

$||...||$ in (2.27) can be replaced by any distance measure. If norm-2 is used, the problem is equivalent to

$$\begin{array}{rl} argmin & \frac{1}{2}||w||^2 \\ s.t. & y_i(x^T w - b_0) \geq 1, \;\; \forall i \end{array}$$

Equation (2.28) is a quadratic programming and according to the strong duality theorem it can be converted to:

$$\begin{array}{rl} argmax & \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j \\ s.t. & \alpha_i \geq 0, \;\; \forall i \end{array}$$

which is called the dual problem of (2.28).In practice, we always use (2.29) as it is easier to handle numerically. Moreover, in (2.29) all the computation of x i is written in terms of inner product, and that means it can be generalized to a nonlinear case by employing Kernel technique.

2.4 Face Database

"Because of its non rigidity and complex three-dimensional structure, the appearance of a face is affected by a large number of factors including identity, face pose, illumination, facial expression, age, occlusion, and facial hair. The development of algorithms robust to these variations requires databases of sufficient size that include carefully controlled variations of these factors. Furthermore, common databases are necessary to comparatively evaluate algorithms. Collecting a high quality database is a resource-intensive task: but the availability of public face databases is important for the advancement of the field" [35]. In this section we briefly review some publicly available databases for face recognition, face detection, and facial expression analysis, and we'll mainly focus on the three databases which we will use in this thesis.

To facilitate this statement, we divide face databases into two categories according to their designing goals. In the first part, we'll introduce databases which are normally used for face recognition; those which are dedicated to expression recognition will be discussed in the second part. As only a few databases are of the second type, and FER system shares some common modules with identity recognition system, in this work we also use some databases of the first type.

2.4.1 Databases For Identity Recognition

Most face databases are of this category (Table.2.1). To test for robustness, some of them are captured under different poses, illuminations and expressions. However, because they're mainly designed for identity recognition, the expressions are added as noise and usually not well controlled. So in general these databases are considered not suitable for FER research. In our work, we only use them to train peripheral modules (processing and Feature Extraction).

The IMM Face Database [66]

The IMM Face Database comprises 240 still images of 40 individuals (7 females and 33 males), all without glasses. For each person, 6 images are provided:

Table 2.1: Some of the most popular Face Recognition Databases [35]

Database	No. of subjects	Pose	Illumination	Facial Expressions
AR	116	1	4	4
BANCA	208	1	++	1
CAS-PEAL	66-1040	21	9-15	6
CMU HYPER	54	1	4	1
CMU PIE	54	1	4	1
Equinox IR	91	1	3	3
FERET	1199	9-20	2	2
Harvard RL	10	1	77-84	1
IMM FACE	40	3	2	3+
KFDB	1000	7	16	5
MIT	15	3	3	1
MPI	200	3	3	1
ND HID	300+	1	3	2
NIST MID	1573	2	1	++
ORL	10	1	++	++
UMIST	20	++	1	++
U.Texas	284	++	1	++
U. Oulu	125	1	16	1
XM2VTS	295	++	1	++
Yale	15	1	3	6
Yale B	10	9	64	1

- Frontal face, neutral expression, diffuse light.
- Frontal face, happy expression, diffuse light.
- Face rotated approx. 30 degrees to the persons right, neutral expression, diffuse light.
- Face rotated approx. 30 degrees to the persons left, neutral expression, diffuse light.

- Frontal face, neutral expression, spot light added at the persons left side.

- Frontal face, joker image (arbitrary expression), diffuse light.

The images are stored in 640 480 JPEG files. Owing to technique problems, most images are RGB, but some are grey-scale [66]. One good thing about this database is that manually labeled face contour is available. The following facial structures were annotated using 58 landmarks: eyebrows, eyes, nose, mouth and jaw. These landmarks are divided into seven point paths; three closed and four open as shown in Fig.2.14. In our work, this database will be used to train the ASM and AAM model.

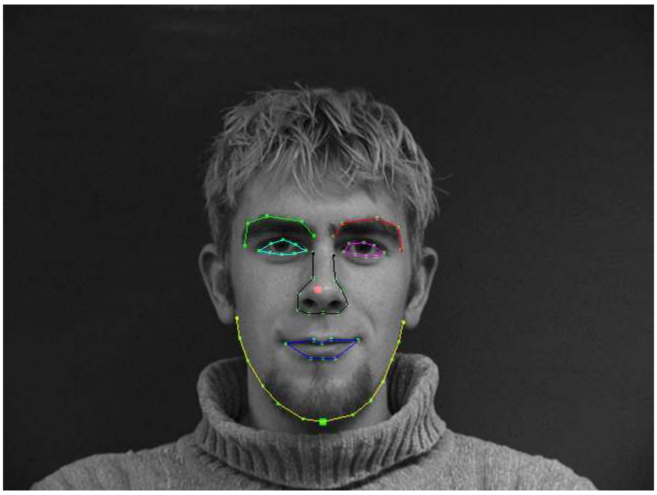

Figure 2.7: Example image from IMM face database

CMU Pose, Illumination, and Expression Database [79]

The CMU-PIE database is among the most comprehensive databases in this area. It systematically samples a large number of pose and illumination conditions along with a variety of facial expressions. The PIE database was captured under 21 illuminations (lit by 21 flashes) from 13 directions (using 13 synchronized cameras). In total, there are 41,368 images obtained from 68 individuals. In our experiment, we only use a sub-set of this database which consists of images of 62 people. 25 images were selected for each individual with 5 different viewpoints and 5 different illuminations. Part of the data set is shown in Fig.2.15.

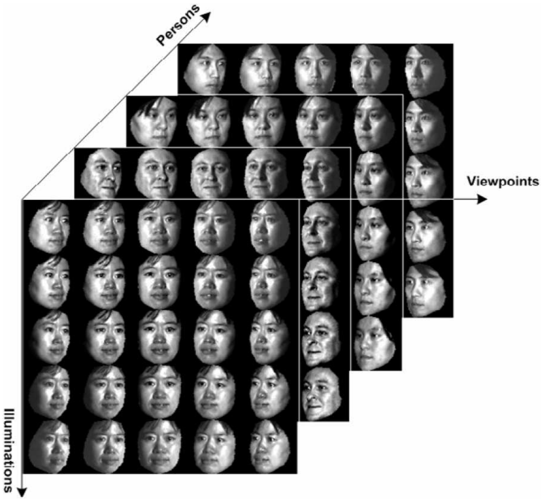

Figure 2.8: A subset of CMU PIE database [53]

2.4.2 Databases for Expression Recognition

"The human face is able to display an astonishing variety of expressions. Collecting a database that samples this space in a meaningful way is a difficult task" [35]. As a result, there are many fewer databases available for expression recognition (Table 6.1). As mentioned in 2.1.1, there are two ways to describe facial expressions. Available databases can be categorized into two classes according to the description they used. In one group [38] expressions are coded in FACS, while in the other group [57] images are labeled by their sprototypic emotional expressions.

Japanese Female Facial Expression Database [57]

The JAFFE database contains 213 images of 10 Japanese female models. Their images are labeled by emotions: six basic emotions (anger, disgust, fear, joy, happy, sad and surprise) are considered and Neutral is added as the 7th emotion which is defined through the absence of expression. Fig.2.16 shows example images for one subject along with emotion

Table 2.2: Commonly used expression recognition databases [35]

Database	No. of subjects	No. of Expressions	Image Resolution	Video/Image
JAFFE	10	7	256 X 256	Image
U. Maryland	40	6	560 X 240	Video
Cohn-Kanade	100	23	640 X 480	Video

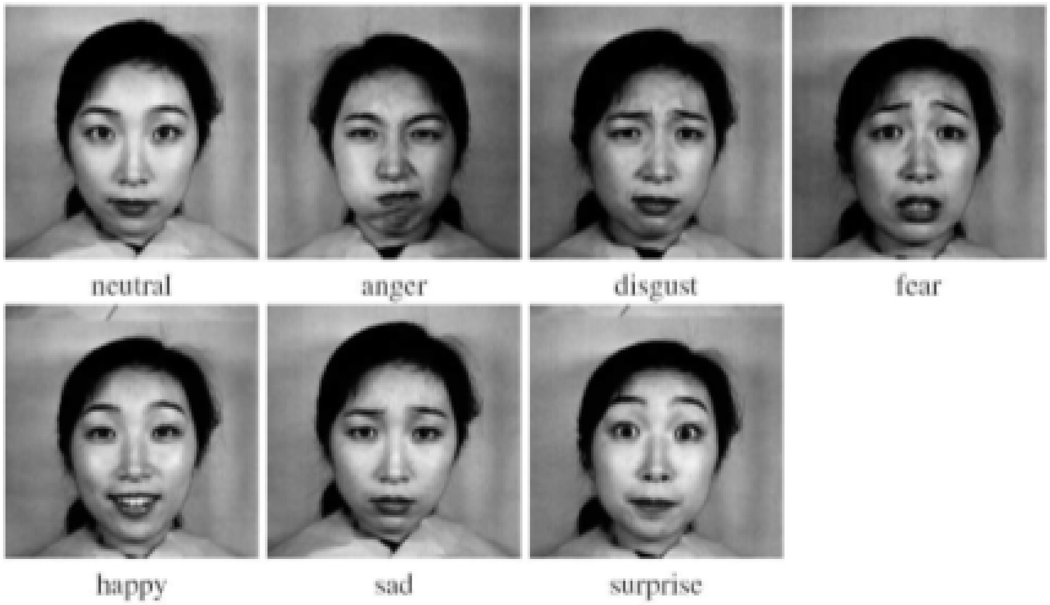

Figure 2.9: Example images from JAFFE database[35]

labels. The images were originally printed in monochrome and then digitized using a flatbed scanner.

2.5 Chapter Summary

In this chapter, we first talked about the background of facial analysis, then gave an overview of the development in this area, and we also briefly introduced some state of the art techniques which might be useful for our system. At the end, we had a glance at some face databases for identity and expression recognition. Starting in the next chapter, we'll discuss the design of our FER system.