Chapter 1

Introduction

Facial expression can convey a wealth of information. From gross emotional state to vaguer but still semantically meaningful structures. The configuration and movement of facial features offers valuable insight into a person. The complicated nature of facial expression (there are over 20 muscles dedicated to facial expression, and their movement is subtle and combines non-linearly) has prevented exhaustive statistical studies of facial dynamics.

There is often the capacity to record huge amounts of data related to a particular problem or study. However, while there is good reason to believe that information of interest is hidden in this raw captured data, it is often unclear how to extract relevant information. The size and complexity of the database make human analysis impossible, or at the very least, overly time consuming and costly.

Machine learning can offer a solution to this problem when traditional statistics are not enough. Even though our capacity to record data still greatly surpasses our ability to extract information from it, the field has been growing in importance especially over the past two decades. The current demand in the industry for data mining and machine learning knowledge has resulted in a growing number of scholarships and grants awarded in this area. It is a relatively young field with many open problems, but also others on which important achievements have been done: speech recognition is considered a more or less solved problem (there are a large number of commercial applications that work satisfactorily [25]), and there have been great advances in the field of face recognition that show an at least promising performance in the most advanced systems. Automated facial expression analysis is perhaps the next step.

Facial expression has been of particular interest to the psychological community, as expression has been found to be intimately coupled with emotion. As an example, recent studies have found that divorce can be predicted by the behavior of men: the more they lower the lip corners, the more likely it is they ask for a divorce.

While the first studies published are from the late seventies, it is not until the nineties that practical results in face analysis are possible. The advances in computer vision, especially in face recognition and tracking, and the much increased computing power made a series of studies possible; the problem, though, is still far from satisfactorily solved [17].

1.1 Motivation

The interest on the study of facial expression has grown over the last two decades, perhaps because advances in machine learning, computer vision, and efforts in data collection have made automatic studies of facial expression in video sequences increasingly possible.

Recent studies have begun to seriously explore the applicability of machine learning and computer vision to facial activity studies, with a broad field of possible applications: depression, deception, pain, effects of drugs, driver alertness, etc... [3, 27, 21, 15].

Depression is, with no doubt, one of the biggest problems of the contemporary society. According to the World Health Organization it is estimated that 5-10% of the population at any given time is suffering from identifiable depression and need psychiatric treatment or psychosocial intervention. [19]

The Hamilton test is broadly used for depression assessment and is combined with studies of several illnesses not necessarily related with psychology, such as Alzheimer's disease or fibromyalgia [16, 13]. It consists in interviews between a potential patient and a therapist, who follows a 17 or 25 questions questionnaire. Although this questionnaire is normalized, the depression severity of the patient is reflected in the score the therapist gives to each of the questions and therefore it lacks complete objectivity.

An automatic depression assessment process, based on video/audio sequences of the subjects would be a breakthrough in both the field of psychology and computer vision. On one hand, it would facilitate the task of the therapists, who could complement their decisions with that of the automatic assessment. On the other hand, it will prove that it is feasible to understand the human behavior automatically applying computer vision and machine learning techniques and it would lead the development of methods for the automatic assessment of other illnesses, such as autism.

This project is a first approximation of an automatic depression assessment software based on a two-class classification procedure, being the two classes the *depressed* and the *non-depressed*. Since the classification experiments will be focused on specific parts of the face (e.g. eyebrows, eyes, mouth), it will be possible to determine which of these parts are more discriminative for the depression assessment. Therefore, it can be said that other motivation of this project relies on a better understanding of the facial expressions and human behavior related to depression.

1.2 Objectives

The Spectrum database is used in this project as a data source. It consists in a collection of video sequences from depressed patients (for more details, see Section 3.2), focused in the person's head.

A face tracking method based on Active Appearance Models (AAMs) is applied to each of the video sessions, from which a set of facial features is extracted. Once this features are extracted, after a data reduction step, they are used as input of Support Vector Machines (SVMs) classifiers. This step leads us to the final aim of the project: to automatically determine if a person is depressed or non depressed. In other words, the aim of this project is to build a first approximation of an automatic depression assessment software, able to classify depressed and non depressed people applying state-of-the-art computer vision and machine learning techniques.

It is easy to think that the difficult part of this work is to build a classifier which enables us to separate the *depressed* class from the *non depressed* class. However, this is not the case. The main difficulty in which this project relies is finding the features of the face which are relevant to the depression assessment. For this purpose, previous research in human behavior is taken into account, but also some self intuition, which is essential in order to come up with new theories and algorithms.

It is worth saying that this project does not aim either to improve the computer vision and machine learning methods applied (e.g. AAM tracking, SVM classification) or to invent new ones. It will just exploit the state-of-the-art techniques and adapt them to our specific purposes.

1.3 Outline of the project

This report is structured in three main parts: data, classification and results. Before going into this points in depth, Chapter 2 focuses on giving a brief introduction to the essential depression concepts, such as its definition, the depression assessment process and the severity criterion.

1.3.1 Data

The start point of the data used in this project is the Spectrum database. It is comprised by a set of video sequences in which depressed patients appear while answering the Hamilton test in an interview with a therapist (it is explained in detail in Section 3.2). However, having the right data does not suffice. A state-of-the-art tracker based on Active Appearance Models is applied to track the faces of the depressed patients, which enables to extract the facial behavior information. Finally, since the tracking is not perfect (normally due to occlusions of the face), once all the video sessions are tracked, a clean up step is needed, eventually obtaining the ready-to-use-data.

1.3.2 Classification

The classification process is detailed in Chapters 4 and 5. Chapter 4 presents the maths behind the classifier chosen, the Support Vector Machine, and the performance measures used to determine the goodness of the classification experiments. Furthermore, some normalization methods, understood as the alignment of the faces prior to the feature extraction step, are compared.

In Chapter 5 the feature extraction procedure is detailed. Three layers of features are defined: the feature level, the group level and the global level. This layering will result in a data reduction, an essential step due to the huge amount of data we are dealing with in this project. Besides, a new kind of features will be presented: the *shape coefficients*. They rely on a Principal Component Analysis and are able to represent the behavior of the face as a whole.

1.3.3 Results

The classification experiments will be performed using two kind of features: the landmark points resulting from the AAM tracking, and the shape coefficients. At the same time, two different classification algorithms are used, both based on a K-fold cross validation method. The first one performs a leave-one-subject-out, while the other is a 4-fold. Both techniques are slightly modified in order to deal with the fact that the Spectrum database contains several sessions for the same subject, and including sessions of the same subject in the training and testing sets results in including highly correlated data in both sets.

The performance of the experiments will be given by Receiver Operating Characteristic (ROC) curves, as well as by scalar values such as the F parameter, the accuracy and the area under the ROC curve.

1.4 Framework of the project

This project has been developed at the Human Sensing Lab, a computer vision and machine learning research laboratory part of the Robotics Institute of Carnegie Mellon University, leaded by Dr. Fernando de la Torre, in collaboration with the Affect Analysis Group of the University of Pittsburgh, whose head is the psychologist Dr. Jeffrey Cohn. Since this is an interdisciplinary project that applies computer vision and machine learning techniques in the field of psychology, it could not have been possible without the continuous interaction between these two research groups.

The software developed along the development of this project is coded in MATLAB. MATLAB is a numerical computing environment that allows easy matrix manipulation, plotting of functions and data, and implementation of algorithms. The main advantage of MATLAB is that it is a very high level programming language which makes it very faster to program, but it has the drawback of having an execution speech much slower than C or C++. Since the software developed for this project does not have any restrictions on the velocity of execution (it does not aim to be real-time), MATLAB seems to be a perfect choice.