Emotion Detection using Automatic Facial Recognition

Jonathan Hicks

October 3, 2016

Abstract

This paper covers the basics of facial recognition and discusses a few different methods for facial recognition. The paper then moves towards methods for feature extraction and emotion detection where recent work is the focal point. The end goal of the research within this paper is to produce a program that will take in an image as input, scan for a face, and determine the emotion that face is portraying. To accomplish this, thresholding methods and geometric facial features will be used. The future work section provides more detail on the goal of this research.

1 Introduction

Facial recognition is sub-field of artificial intelligence that is all over the world and recognizing facial expressions is something that human beings have been doing since the beginning of time. It's amazing considering how each and every face is unique. Even the faces of identical twins are unique in some way. People can easily recognize another individual's facial expression based on the positioning of each facial feature (eyes, eyebrows, mouth, nose, etc..). For example, if the eyebrows are raised that person might be surprised. Or if the corners of the mouth are higher than normal that person is most likely smiling and therefore, happy. These are the kinds of things that people and automatic facial recognition systems use to identify facial expressions.

Automatic facial recognition systems are useful in multiple applications. These applications range from scanning your face into video games to helping police identify criminals in surveillance systems to security cameras at ATM machines. There are many more applications out there but the point is that these automatic facial recognition systems are everywhere.

This paper's main goal is to survey the field of facial recognition along with the detection and recognition of human emotion based on facial features. This paper will highlight some common problems in facial recognition and will also provide different algorithms that are used for automatic facial recognition and emotion detection.

2 Detection Process

In this section the objective is to simplify the facial recognition process into steps and then expand into emotion detection. First, the system must be able to identify whether or not a face is present in the image or video. Once a potential face has been located, the system then has to be able to scan that area and compare it with a database to determine if there is in fact a face there. After the face had been confirmed we can then start to look at facial features and use that information to determine the facial expression and emotion of that face.

2.1 Facial Representation

How the faces in a facial recognition system are represented is crucial to the accuracy of that system. The representations for the faces should be stored in a data set. This database is then compared to the image that is scanned by the system to determine if the image is a face. Creating a data set is especially important when trying to determine facial expressions or emotion. For example, Curtis Padgett and Garrison Cottrell from the Department of Computer Science at the University of California conducted and experiment to identify human emotion with facial recognition software[1]. Their original data set included pictures of undergraduate portraying different emotions. However, there were differences with the faces portraying an emotion and the faces feigning that emotion. Those differences resulted in emotion detection being very inaccurate. To correct the issue a validated data set was created by trained actors who were able to accurately portray the emotions. Clearly, being consistent with your data sets is important.

2.2 Facial Detection

Determining whether or not a face is present is perhaps the most important step in the facial recognition process. Mainly because it is essentially the beginning of the first step in the process (besides preparing the data set). Once the data set is ready, some common problems have to be taken into consideration[2].

One common problem is a partially blocked face. When someone sees a picture of someone and that person's face is partially blocked, that person is able to fill in the missing pieces. Not all systems are capable of filling in those pieces. Another common issue is lighting. Again, this is not as big of an issue for people as it is for machines. When the lighting is too bright, people can squint and still make out that a face is present. This is much more difficult for machines.

2.3 Feature Extraction

There have been many different ways for extracting facial features. Zhang et al. [3] used a multimodal learning technique that focused on learning the representations of texture and landmark modality. The texture modality is a collection of image patches highlighting facial key points. The landmark modality depicts the facial key points in a facial expression sequence. This multimodal learning algorithm is essentially an artificial neural network (ANN) where it takes in numbers and modality as input, stacking these inputs in a hidden layer, and outputting a result. Figure 1 below shows what this multimodal algorithm looks like.

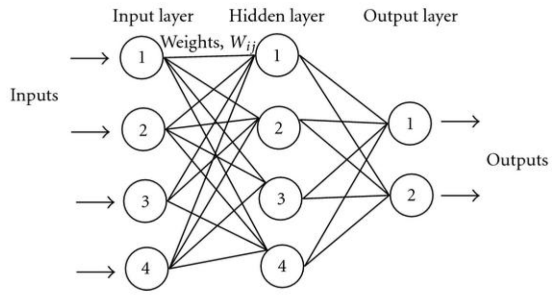

Figure 1: Structure of multimodal algorithm

2.4 Facial Points

The most common method for emotion detection is facial point detection. Facial point detection focuses on specific places on the face such as the corner of the eye, corners of the mouth, tip of the nose, beginning and end of the eyebrows, and there are many different models used in facial point detection. Facial point detection has been extensively studied in recent years. There are two general approaches to facial point detection. The first is to classify search windows and the second is to directly predict the positions of the key points[4]. For classifying search windows, a component detector is trained for each facial point and the window location is based on a local regression. By directly predicting the positions of the facial points, scanning is eliminated and thus, this is more efficient. Sun et al.[4] propose a cascaded regression for facial point detection that include three levels of convolutional networks.

Convolutional networks are neural networks (CNNs) that are similar to artificial neural networks (ANNs) in that they have trainable weights and these weights take in some input. Figure 2 shows an example. The convolutional network of Sun et al. takes in the entire face as input and use texture context information to extract features in deeper structures of the face. The first level of their network makes accurate predictions rather than making a rough estimate. The other two levels are for refining the earlier estimations of the facial points. These two levels are more shallow because they focus on smaller regions of the facial points whereas the first layer is scanning the entire face. At all three levels, multiple convolutional networks are combined to improve accuracy and reliability.

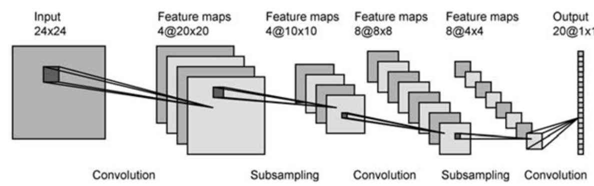

Figure 2: Example of a convolutional network

3 Emotion Detection

The key to accurately depicting emotion from images is to first extract the necessary facial features which are then either applied to different action units (AUs) or directly to the classifiers. One method for emotion detection is to first find the geometrics of certain facial points. This allows us to determine the general shape and location of each facial feature. The following subsections break down each facial feature and use different methods and thresholding algorithms to determine the location of each facial point for that feature.

3.1 Eyes

The eyes are an extremely important facial feature as they help align the face and allow for a good reference point to other facial features. In one study, Majumder et al. detect the eyes and then propose a thresholding algorithm that takes in input from every pixel from the eye region[5]. This algorithm is represented as

$$T_{local}(x, y) = \mu_{global}(x, y) + k * \sigma_{local}(x, y)$$

$$\mu_{global}(x, y) = (1/M * N)[\sum_{j=0}^{N} \sum_{i=0}^{M} f(i, j)]$$

where $T_{local}(x, y)$ is the threshold value of the pixel at (x, y) and $\sigma_{local}(x, y)$ is the standard deviation. μ_{local} is the local mean and μ_{global} is the global mean and $M * N$ is the size of the eye.

3.2 Eyebrows

The location of the eyebrow is first found using facial geometry. Once located, the facial points are stored in a vector. The positioning of these facial points will aid in determining the emotion.

3.3 Nose

As the nose lies between the eyes and represents the approximate center of the face, the rough estimation of the nose should be simple. The thresholding algorithm from earlier can now be applied to the nose as well as a contouring method to find the two nostrils.

3.4 Lips

Majumder et al.[5] provide a step-by-step algorithm for estimating the location of the lips. The first step is to find the center of the eyes and the height of the nose. From there they form a rectangle around the projected region of the mouth by using the function $rect(x_I, y_I, h_I, w_I)$ where x_I and y_I are the coordinates of the top left corner, h_I is the height and w_I is the width.

4 Future Work

Going forward I hope to gather more research and more information about facial recognition and emotion detection. Once I have obtained all the information that I need, I plan to use opencv and its libraries to develop software that will scan an image for a face and, if there is a face, determine the emotion that particular face is portraying.Opencv has source code on the website that I can use to find the facial points of faces in images. I will then be able to add code to that existing code to determine emotion based on the location of the facial points. This added code will consist of using geometric facial features and thresholding algorithms as described in section 3. Opencv also has a validated database of images that I can download and use to compare with the images that my software will scan.

References

[1] C. Padgett and G. Cottrell, "Representing face images for emotion classification," p. 895900.

[2] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, no. 1, p. 6577, 1992.

[3] W. Zhang, Y. Zhang, L. Ma, J. Guan, and S. Gong, "Multimodal learning for facial expression recognition," *Pattern Recognition*, vol. 48, no. 10, p. 31913202, 2015.

[4] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[5] A. Majumder, L. Behera, and V. K. Subramanian, "Emotion recognition from geometric facial features using self-organizing map," *Pattern Recognition*, vol. 47, no. 3, p. 12821293, 2014.