

Automatic Human Emotion Detection

Retshidisitswe Lehata

Thesis presented in fulfilment
of the requirements for the degree of
Bachelor of Science Honours
at the University of the Western Cape

Supervisor: Mehrdad Ghaziasgar
Co-supervisor: Reg Dodds

This version April 18, 2017

Declaration

I, RETSHIDISITSWE LEHATA, declare that this thesis “*Automatic Human Emotion Detection*” is my own work, that it has not been submitted before for any degree or assessment at any other university, and that all the sources I have used or quoted have been indicated and acknowledged by means of complete references.

Signature:

Date:

RETSHIDISITSWE LEHATA.

Abstract

Applying Facial expression recognition to understand customer satisfaction, during face-to-face interactions between an employee and a customer. The Viola Jones Algorithm is used to detect the face and the Pyramid Histogram of Oriented Gradients(PHOG) to extract the features from the face. Lastly the training is done using Artificial Neural Networks. . .

Acknowledgment

This thesis is a compilation of the efforts of many people that helped and me through the years. I would first like to thank my supervisor Mehrdad Ghaziasgar encouraging me during my study. Without our weekly meetings, this work would not have been possible.

Contents

Declaration	iii
Abstract	v
Acknowledgment	vi
List of Figures	viii
1. Introduction	1
1.1 Problem Statement	1
1.2 Proposed Solution	1
1.3 Proposed Method	2
2. Related Work	4
2.1 Capture Image from the camera and image processing	4
2.2 Face Detection - finding the face	5
2.3 Feature Extraction- extract the features from the face	6
2.4 Train machine learning technique	7
2.5 Results	8
Bibliography	9

List of Figures

2.1	4 stages for facial expression recognition	4
2.2	Histogram Equalization	5
2.3	Preprocessed image with Oval Masks	5
2.4	Architecture of an artificial neuron and a multilayered neural network	7

Chapter 1

Introduction

Communication plays a large role, in daily human interaction. It can take the form of verbal or non-verbal communication, Mehrabian found that over 93% of verbal communication is conveyed through, the tone of the voice (38%) and non-verbal cues(55%)[1]. Understanding non-verbal communication is a valuable skill, in that it is a universal form of communication. Non-verbal communication is a combination of body language, physical gestures and facial expressions. Ekman & Friesen found six facial expressions that are universally identifiable in recognizing Fear, Anger, Disgust, Surprise, Happiness and Sadness[2]. A facial expression is made up of the changes in facial muscles (mouth, eyes, eyebrows etc.), these changes reflect ones current state of mind.

1.1 Problem Statement

Companies use feedback from their customers as a metric to measure their customer satisfaction rate. Customer feedback can be initiated by the customer as a compliment or criticism, based on the service they were given. Alternatively, a company can offer their customers optional online or physical surveys to be completed. The problem is that customers are more likely to refrain from commenting on a service, unless provoked to do so.

1.2 Proposed Solution

Customer service is a large revenue stream for companies whose core business is based around the customer. Ensuring that the customer stays happy is key, for their success. An Automatic Human Emotion Detection (AHED) system can be applied in any environment that can benefit from understanding facial expressions. The proposed solution combines customer satisfaction with an automated system. Using face detection and facial feature extraction to

identify the dominant customer emotion. The training and classification of the system will be done using a machine learning technique. Companies can incorporate the results of the system in improving their customer service, ensuring that their customers stay satisfied.

1.3 Proposed Method

The proposed method for this project, first captures an image from the camera, the image is processed using the Viola Jones Algorithm to detect the face and the Histogram of Oriented Gradients(HOG) to extract the features from the face. Lastly the AHED system is trained using Artificial Neural Networks, to classify each emotion.

Chapter 2

Related Work

There are a wide variety of methods that are used in the field of emotion recognition using facial expression, however a large number of those methods are implemented using a similar process. Initially the image is captured and processed, thereafter the face is detected and features are extracted. Then a learning technique is trained to do the classification of the emotions. The related work looks at all 4 aspects of the implementation process and different methods used by other researchers.

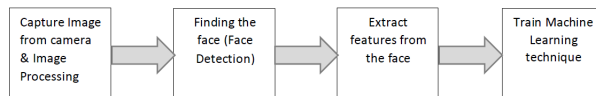


Figure 2.1: 4 stages for facial expression recognition

2.1 Capture Image from the camera and image processing

An image is taken from the camera and image processing tools are used to help normalize the input image.

- Goyal and Mittal, achieved the desired resolution and colour for their images by adjusting the brightness and contrast of the image[3].
- Reddy and Srinivas, scaled and cropped their images to (150×120) , and ensured that the location of the eyes was the same in each image. The image is further processed, using an average combination of all the input image histograms. This process is called histogram equalization and helps in decreasing variation in an image. Histogram equalization is a

technique for stretching out the intensity range of an image to enhance the contrast of the image[4].

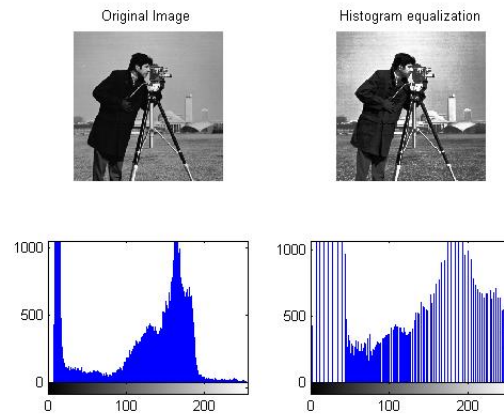


Figure 2.2: Histogram Equalization

- Boubenna and Lee, scaled their images to (100×100) pixels[5].

2.2 Face Detection - finding the face

Finding the location of the face helps identify the region that contains all the features required to continue with facial expression recognition. The rest of the image is not as important for this purpose.

- Reddy and Srinivas, applied a fixed oval shaped mask over the image to extract the face region[?].

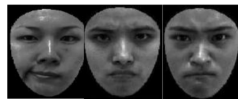


Figure 2.3: Preprocessed image with Oval Masks

- Boubenna and Lee, used the Viola and Jones algorithm to detect the location of the face in an image. [6]This algorithm uses Haar-like features to help find the facial features, such as the eyes , nose and mouth. The Ada Boost algorithm is used to reduce the number of features, if there are too many. They used the canny edge detection operator to detect the edges of the face.[5]

2.3 Feature Extraction- extract the features from the face

Once the face has been detected, it is important to identify which features will be used for feature extraction (eyes, nose, mouth, eyebrows, full-face etc.). The feature extraction algorithm can be applied based on its compatibility with the features chosen.

- Goyal and Mittal, extracted the nose, mouth and eyes using the Viola and Jones Haar classifier[3].
- Reddy and Srinivas, considered the entire face for the feature extraction not just the eyes, mouth and nose individually . First, they used Gabor filters to generate a bank of filters at 5 spatially varying frequencies and 8 orientations. The filtered outputs were then concatenated and used PCA(Principle Component Analysis) to reduce dimensionality. The PCA algorithm will generate the eigenfaces for each image of dimension($N \times N$). From this their system generated the eigenvector of dimension ($N/2$) for each image. PCA is a statistical technique that reduces the dimensions of feature vectors. The high dimensionality of feature vectors can cause over-fitting during classification. The vectors that relay the distribution of the face images the best are selected, and are used to define the subspace(face space) of the face images[4].
- Boubenna and Lee, used Pyramid Histogram of Oriented Gradients(PHOG) to extract features. PHOG represents an image by its local shape and the spatial layout. The local shape of an image is represented by a histogram of edge orientations within an image sub-region, which are divided into K bins. The Spatial layout is represented by tiling the image into regions at different levels. Each image is divided into a sequence of increasingly finer spatial grids by repeatedly doubling the number of divisions in each axis direction[7]. The parameters of PHOG were set as follows: 3 for number of level, 360 degree for angle and 16 for number of bins. To decrease the number of features a Genetic Algorithm(GA) is used, which resembles natural selection to find the optimal features[5].

2.4 Train machine learning technique

The training of the machine learning technique based on supervised learning. Where the machine learning technique is given labelled images (Happy, Sad, Anger, Disgust, Surprise, Fear and Neutral) and is required to learn them. Once the machine learning technique has completed its training it can then be fed unlabelled images, and the result would be a prediction of which label best suits the given image.

- Goyal and Mittal, used an Artificial Neural Network, with one hidden layer. The neural network architecture has three types layers: input, hidden and output layers. Feed-forward ANNs allows the signal to travel in one direction from the input layer to the output layer. Recurrent networks contain feedback connections, where the signal moves in both directions. To get accurate results from the ANN, the weights can be set explicitly using prior knowledge or the ANN can be trained to help find the optimal weights[3, 8].

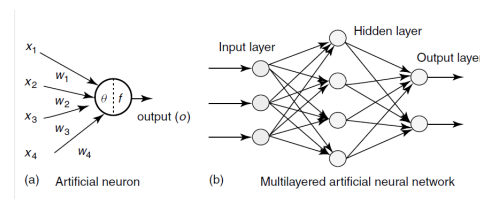


Figure 2.4: Architecture of an artificial neuron and a multilayered neural network

- Reddy and Srinivas, used an Artificial Neural Network, with two hidden layers[4].
- Boubenna and Lee, used Linear Discriminant Analysis(LDA) and K Nearest Neighbours(KNN). LDA finds the maximum distances within classes, to obtain maximum class separation. LDA only uses up to second order moments, such as the covariance and mean, of the class distribution. KNN classifies unlabelled samples according to the training samples. KNN finds the nearest K in the labelled samples and set them to the closest group, for the unlabelled samples. One distance measure is required, and [5] used the cosine distance measure.

2.5 Results

The results from the three studies are as follows, with [5] having the best overall results for their human facial expression system.

- Goyal and Mittal, achieved a 80% Right classification, using a confusion matrix and a regression plot to verify the results[3].
- Reddy and Srinivas, achieved a 85.7% Right classification using the JAFFE database[4].
- Boubenna and Lee, achieved a 99.33% accuracy, using the Radboud Faces Database(RaFD)[5].

Bibliography

- [1] A. Mehrabian. *Nonverbal communication*. CURRENT CONTENT, CA, USA, 1984.
- [2] P. Ekman. *NONNERBAL BEHAVIOR AND COMMUNICATION*. Lawrence Erlbaum Association, New Jersey, 1977.
- [3] R. Goyal and T. Mittal. Facial expression recognition using artificial neural network. *HCTL Open Int. J. of Technology Innovations and Research*, 10:1–10, July 2014.
- [4] Ch. Satyananda Reddy and T. Srinivas. Improving the classification accuracy of emotion recognition using facial expressions. *International Journal of Applied Engineering Research*, 11(1):650–655, 2016.
- [5] H. Boubenna and D. Lee. Feature selection for facial emotion recognition based on genetic algorithm. *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, 12:511–517, August 2016.
- [6] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 2(57):137–154, 2004.
- [7] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. *CVIR*, pages 401–408, July 2007.
- [8] A. Abraham. *Handbook of Measuring System Design*. John Wiley & Sons, Ltd., OK, USA, 2005.