# Face expression recognition for social good
# - smart interactions matter -

– ITSG report –

**Author**
Bogdan-Daniel Bălănescu
Software Engineering, 258-1

2019

**Abstract**

This paper studies the problem of Face Expression Recognition (REC) in an attempt to build a tool that helps novice actors better analyze their performance and get real-time feedback to improve. [The work for the next laboratories will add here: short presentation about the intelligent method used for solving the problem in this paper, data involved in the numerical experiments, comparison with other state of the art researches and conclusion with the obtained results].

# Contents

# Chapter 1

# Introduction

## 1.1 What? Why? How?

For novice actors, it is often hard to control their emotions and to express what is on their script well, so the need for an expert to be with them when they rehearse is needed. A tool that would give them feedback in real-time may just make their start easier. It could be there for them anytime and for free, rather than hiring an expert to watch them rehearse.

This paper addresses this problem by proposing a tool able to recognize emotions from real-time video footage or photos.

- The problem of Face Expression Recognition (FER) is a classification problem which has received an important amount of attention in the last decade with various approaches such as the feature-based Tree-Augmented-Naive Bayes (TAN) classifier, Local Binary Patterns (LBP) classifier, Support Vector Machines and numerous Neural Networks based approaches.

- The importance of FER is present in a variety of domains, such as: psychology, neuroscience and even philosophy. Science today still cannot explain for sure where do emotions come from or if emotions are the ones driving our decisions. We will focus on self-improvement, providing a tool capable of helping people realize if they have healthy relationships or not.

- In this paper, we present an artificially intelligent solution to recognizing emotions from pictures. [At this point it is only speculation] We will take a dataset of pictures labeled with emotions and based on a number of facial points (features) we will train a model able to classify emotions.

[The work of next laboratories will add here: A short discussion of how it fits into related work in the area. Summarize the basic results and conclusions that we will present.]

## 1.2   Paper structure and original contribution(s)

[Should present here this information when several tested models are acquired: The research presented in this paper advances the theory, design, and implementation of several particular models. (if it is the case)]

The main contribution of this report is to present an intelligent algorithm for solving the problem of Face Expression Recognition.

The second contribution of this report consists of building an intuitive, easy-to-use and user friendly software application. Our aim is to build an application that will help novice actors to better analyze their facial expressions and get real-time feedback while rehearsing. Features present in the application:

- The user shall be able to take a picture and see the emotions of the people present in the picture (labeled on the picture).

- The user shall be able to import an already existing picture into the application and see the emotions of the people present in the picture (labeled on the picture).

- The user shall (maybe) be able to record a session of his interactions with others and see the video with the emotions of the people present in it (labeled while streaming the video).

The third contribution of this thesis consists of providing a comparison between state of the art results in FER and the proposed algorithm.

[Should be present in the next laboratories: The present work contains $xyz$ bibliographical references and is structured in five chapters as follows.]

In the second chapter we will take a look at a formal introduction of our FER problem and weight the advantages and disadvantages of using AI to solve it.

The second chapter will describe the state of the art in FER.

In the the third chapter we will further detail our proposed approach to solving the problem.

[Short placeholder here, until we reach detailing the third chapter: (dataset with pictures and labels (emotion) -> new dataset facial points (features) and labels (emotion) -> model to solve the problem)]

The fourth chapter will present a comparison for the chosen methodology, data and results between two different approaches to solving this problem (ours and one more).

The final chapter will present our conclusions and future work to be done regarding this problem and application.

# Chapter 2

# Scientific Problem

## 2.1 Problem definition

For people pursuing their hobby of acting, it might be hard to hire an expert while rehearsing or even coming to terms with the tight schedule of a hard working day and a few minutes of spare time to rehearse a play. The need for a tool that would assist people when rehearsing, giving them feedback about their facial expressions in real life, arises and we pursue to deliver such an application.

The solution to such a tool is required to use an intelligent algorithm, because as far as we know, there exist no other methodologies for approaching this problem and it also falls into the category of complex problems that may be more easily solved using a neural network or other intelligent algorithms.

Advantages of solving the problem with an intelligent algorithm:

- it is much faster to reach a solution than proceeding with finding a non-intelligent algorithm to solve it

- it may be impossible to solve it using non-intelligent algorithms due to the vast lack of knowledge in the area of human emotions

Disadvantages of solving the problem with an intelligent algorithm:

- it may require a lot of work in training and retraining models until we reach a suitable (efficient and accurate) model to solve the problem

- it is impossible to reach, using A.I., a solution which has a 100% accuracy

Short description of our initial approach:

- Take an existing dataset containing:

  - images

  - labels: emotions

- Train a model to determine emotions the provided pictures

- Use that trained model in an application to give real-time feedback about the emotions of the people in the picture (initially with just one person, but may expand it to work with multiple people)

# Chapter 3

# State of art/Related work

## 3.1 State of the Art

In this section we present a few methods utilized in order to solve the Facial Expression Recognition (FER) problem.

First, we will take a look at [3], where the problem is solved using k-NN (Nearest Neighbors) and MLP (Multilayer Perceptron).

- What kind of data did they use? Coefficients describing elements of facial expressions (as features) and a range of seven emotional states (as labels). The emotional states detected are: neutral, joy, sadness, surprise, anger, fear, disgust.

- How does their model(s) work? Using Microsoft Kinect 3D for face modeling, they were able to extract 3D models of the face as 3D points, but Kinect 3D also can extract Action Units (AC) based on those points, which basically represent certain features of the face. Choosing 6 of those Action Units (upper lip raising, jaw lowering, lip stretching, lowering eyebrows, lip corner depressing, outer brow raising) they were able to train a 3-NN and an MLP classifier in order to solve the FER problem.

- What were their results? They tested the models for two cases: a) subject-dependent and b) subject-independent. For the 3-NN classifier, they got around 95-96% accuracy and for the MLP algorithms the results were around 75-76%.

Second, let us look at [2], where three significant challenges in FER are discussed: illumination variation, head pose and subject-dependence. We shall focus on the ones that target visual-only databases.

- What about illumination variation? The paper compares different approaches to FER, like SVM (Support Vector Machines) with 31-50% accuracy, Deep Networks with 48-96% accuracy and KNN with 92-96% accuracy. The most utilized dataset was the CK+ dataset [TODO - provide link]. The paper suggests using Fast Fourier Transform and Contrast Limited Adaptive Histogram Equalization (FFT+CLAHE) to overcome poor lighting conditions, among other techniques.

- What about subject-dependece? Subject-dependence means the model is only able to recognize the expressions of the faces it trained with. The paper proposes a solution, by Zhang et al. [TODO - properly cite this here] where temporal, geometric face features were extracted, and using a part-based hierarchical recurrent neural network (PHRNN) to model the facial morphological variations, a multi-signal convolutional neural network (MSCNN) to find the spatial features of face, an accuracy of around 98.5% can be achieved. The used dataset was CK+.

Last, but not least, let us take a look at [1], a practical approach using a Convolutional Neural Network (CNN).

- What kind of data did they use? The used datasets were MMI and CKP [TODO - provide links for the both of them] and they recognized emotions from the following list: anger, sadness, disgust, happiness, fear and surprise.

- How does their model(s) work? Their proposed model is independent from any other third-party feature extraction frameworks and it performs better than the previously proposed CNN models, with an accuracy of 93-99% on the above mentioned datasets. Their model begins with a convolutional layer over the input and then filtering max pooling layer, before entering two fully connected layers (comprised of convolution + pooling, then convolution + convolution, then concatenation and pooling), after which they classify the images with softmax layer. [TODO - maybe provide a picture from the article + citation)

- What were their results? Their accuracy was around 99%.

So far, we have seen three great state of the art examples. One which uses an third-party framework to manipulate the dataset so that the model can be a very simple one to train, like a 3-NN. Another set of examples where top state of the art models were presented and certain impediments were discussed (such as illumination variation and subject-dependency) and one other example which is independent on any other third-party framework in its learning, while still achieving good results.

## 3.2   Useful Tools

Now, let us give a list of useful tools to use when developing intelligent applications:

- Tensorflow, is a Python framework for building machine learning models. A javascript version, Tensorflow.js, also exists.

- Tensorflow Lite, a framework for building machine learning models compatible with portable devices (i.e. mobile phones).

- ML Kit for Firebase, a mobile SDK that empowers mobile applications with Google's machine learning packages. It is also possible to host one's own machine learning model in Firebase (have not experienced with this yet).

- YOLO: Real-Time Object Detection, is a state of the art object detection system, but which can also be trained for a different purpose. It is open source and written in C++.

- fastAI, is a Python framework for buildign machine learning models. Also comes with pre-trained models for certain problems.

# Chapter 4

# Proposed approach

# Chapter 5

# Application (numerical validation)

# Chapter 6

# Conclusion and future work

# Bibliography

[1] Peter Burkert, Felix Trier, Muhammad Zeshan Afzal, Andreas Dengel, and Marcus Liwicki. *DeXpression: Deep Convolutional Neural Network for Expression Recognition.* 2016.

[2] Najmeh Samadiani, Guangyan Huang, Borui Cai, Wei Luo, Chi-Hung Chi, Yong Xiang, and Jing He. A review on automatic facial expression recognition systems assisted by multimodal sensor data. *Multidisciplinary Digital Publishing Institute (MDPI)*, 2019.

[3] Pawel Tarnowski, Marcin Kolodziej, Andrzej Majkowski, and Remigiusz J. Rak. Emotion recognition using facial expressions. Science Direct, 2017.