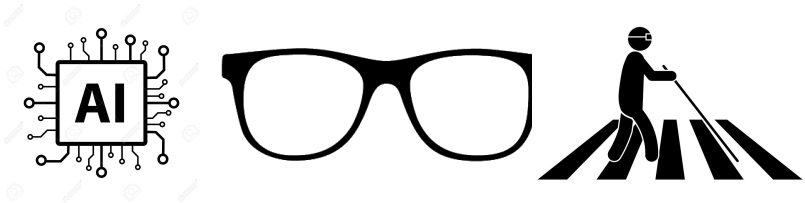
UNIVERSITY OF TWENTE.

Faculty of Applied Mechanics and Data Analysis

A first step towards AI glasses for blind people

Integration of an efficient, real-time gender recognition algorithm into prototype glasses, trained on data reduced by a data reduction algorithm.



H.A. Jekel

B.Sc Thesis
Proposal
Febuary 2021

Supervisors:
prof. dr. ing. B. Rosic

Faculty of Applied Mechanics
and Data Analysis
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Contents

Chapter 1

Introduction

The world as we know it, would not be the same without Artificial Intelligence (AI). It is not unlikely that the world will become progressively more dependent on robots and AI. In 2020, already 26% of jobs are replaced by robots. For this reason, a lot of human skills will become redundant. [1]

In particular, healthcare becomes progressively more dependent on AI. Currently, AI in healthcare includes mostly diagnosis and treatment recommendations, patient engagement, patient adherence, and administrative activities. [2] This thesis will use AI to find a solution to the medical condition blindness.

Visual impairment is a common health problem. According to 2018 statistics, nearly 253 million people in the world are suffering from some form of visual impairment. [3] To help this group of people, this bachelor thesis will make a first step towards a portable face-recognition system to enable blind people to immediately recognize the people in their near vicinity. This is similar to the application ”Seeing AI” [4], which already implements AI to determine gender, hair color and age.

To make this first step, this thesis addresses three main topics. First, neural networks are defined and a description of how they can be used in a gender recognition system is stated, then a gender recognition algorithm is selected based on efficiency and finally a data reduction algorithm is selected that reduces the size of the data needed for training. Additionally, if time allows, a working prototype will be submitted.

This thesis is organized as follows, section 2 discusses the problem statement, containing the problem itself, the goal of the research and the relevance of the research. Next, section 2.1, presents a literature study, containing a first examination of the three main topics of this research. Then, section 4, states research questions and sub-questions. After that, section 5, introduces the methods to be used to answer the research questions. Thereafter, section 5, states the methods to answer the research question and the conditions for these methods. Finally, section 7 displays a table of the planning of this thesis.

Chapter 2

Problem statement

2.1 Problem

In this thesis the research analysis will start with the study of neural networks that can be used for gender recognition. A selection will be made based on efficiency, while keeping accuracy a second aim. The special focus will be paid to deep learning algorithms. Once the algorithm is selected, it will be trained on the collected facial data set to find patterns and rules, obtaining the machine learning model. The training of the neural network will be done on a computer with a higher computational power then that of a Raspberry Pi 3, due to the size of the training data. Currently, large data sets are needed to train the neural network. Training the algorithm will be done using supervised learning, where training samples are labeled. In this assignment the focus will be on reducing the size of the data set for training, which can be done using techniques like singular value decomposition (SVD) [5]. The trained model can then be implemented on the Raspberry pi 3 to make predictions about the gender of the persons in front of an integrated camera. As a last step the prototype will be designed, creating AI glasses for blind people. Using a Raspberry Pi 3 will ensure that the prototype is portable.

2.2 Goal

The ultimate goal is to design a working prototype that determines the gender in real time and communicates this to the blind person, using an efficient gender recognition algorithm and an algorithm, like SVD, that minimizes the training data needed. Additionally, a recommendation for further research will be given. The prototype is optional, and will only be included if time allows.

2.3 Relevance

There are lots of implementations of gender recognition using a Raspberry Pi 3, thus implementation is not necessarily relevant. However, understanding of how to optimize the efficiency of the gender recognition algorithm and understanding how to minimize the training data needed using an algorithm, like SVD, is relevant, even out of the scope of this thesis. In any real-time application of a gender recognition algorithm, it is important to maximize the efficiency of the algorithm to obtain results in real time. Also, minimizing the training data needed will result in less training time, decreasing the costs for training.

Chapter 3

Literature study

3.1 Face detection

To fullfill the goal stated in section 2.2, one has to understand the gender recognition system. A gender recognition system starts with face detection. One way to detect faces is to use Haar Features [6], that are good at detecting edges and lines. The features depicted in Fig. 3.1 are used to recognize the eyes and the nose given an image of a humans face. The features manage to do this as the eyes are generally darker than the cheeks and the nose is generally brighter than the eyes.

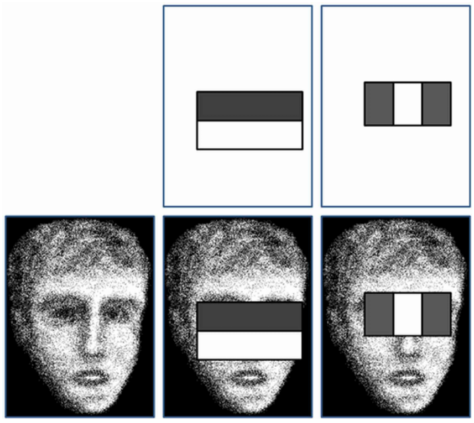


Figure 3.1: Two Haar Features that recognize the eyes and the nose, respectively [6]

However, Haar Features need to be determined manually and for that reason there is a certain limit to the face detection capabilities. Given only edge and line Haar Features, a Haar-based classifier [7] may not be able to recognize a face if it is disturbed (sunglasses, tilted to the side). An advantage of a Haar Feature classifier is that it can be created with a relatively small data set as only the weightings for each feature needs to be determined. Convolutional kernels, unlike Haar Features, are determined by training and have, for that reason, a higher degree of freedom. This makes them capable of recognizing partially covered faces. The tradeoff between convolutional kernels and Haar based features is the speed and accuracy of the face detection. Haar-based classfiers can recognize faces with a higher execution speed, because the Haar-features do not have to be trained, however they can not detect a partially covered or turned face, where the convolutional kernel can. The Haar-feature classifier can be created with a relatively small data set as only the weightings for each feature needs to be determined.

3.2 Neural network

A neural network consists of neurons that use activation functions [5], such as ReLU and sigmoid activation functions. An artificial neural network is formed by placing these neurons in parallel, in series or both. Adding layer upon layer of neurons will result in a deep neural network. There are many different architectures of neural networks, each good at solving a certain problem. The problem of gender recognition is solved using the CNN (convolutional neural network) [8] architecture.

3.3 Post-training pruning

To be able to fit a large neural network on the Raspberry Pi 3, post-training pruning can be used. This technique takes the trained convolutional neural network and examines the influence of each parameter based on their weights on the final value of the network, allowing for real-time inference. [9]

3.4 Machine learning libraries

Google developed a free open source machine learning library called Tensorflow [5]. Keras [5] is a neural network library that can be used on top of Tensorflow to develop a CNN. To determine the prediction capabilities of the CNN the trained model should be evaluated on data that is not present in the training set.

3.5 Singular value decomposition

Singular value decomposition (SVD) [5] is a data reduction algorithm. It can be seen as a data-driven generalisation of the Fourier transform (FFT). Where the FFT uses sines and cosines to approximate functions, to map a system of interest into a coordinate system where the system becomes simple. When systems become complex, there is no obvious coordinate system to map to, to make the system simple. In this case SVD can help, where it takes the data from the specific complex system, and tailors a coordinate transformation based on the data available. SVD can take in high dimensional data and distil it down into the key features, the key correlations. This allows for understanding and modelling of the data.

3.6 Current solutions

As stated in section 1, Seeing AI [4] offers an application that determine gender, hair color and age. However, to use this application, the user has to take out a phone, start an application and then use the application before the phones microphone informs the user. In comparison prototype glasses give instant information about the gender of the person in front of the user. This is a major advantage as information delivery time is reduced.

Chapter 4

Research question and sub-questions

Implementation of gender recognition on a Raspberry Pi 3 has been done before. This thesis, however, focuses on the optimization of the gender recognition algorithm concerning the efficiency as well as optimizing supervised learning techniques to reduce the size of the required data set for training.

4.1 Main research question

- Which specific gender recognition algorithm and which data reduction algorithm should be chosen in order to optimize the implementation of gender recognition using a camera and a Raspberry Pi 3?

4.2 Sub-questions

- What is a neural network and which neural network should be used?
- Which gender recognition algorithm should be selected when focusing on efficiency rather than accuracy?
- Which data reduction algorithm is best for reducing the size of the required data sets for training?

Chapter 5

Methods

This section describes the methodology needed to answer the research questions defined in 4. Once the research questions have been answered, the trained gender recognition model will be implemented on the Raspberry pi 3 to make predictions about the gender of the persons in front of the integrated camera. As a last step the prototype will be designed, creating AI glasses for blind people.

5.1 Neural network

In the first phase of the thesis, a literature study will be conducted to find the most appropriate CNN to classify the gender, based on efficiency and to find a data reduction algorithm to minimize the size of the training data. Next, the model will be subjected to different data sets, containing human faces. The data set that gives the model the highest prediction capabilities, when run on the Raspberry Pi 3 with integrated camera, will be selected.

5.2 Hardware and Programming

The Lenovo P50 with a NVIDIA Quadro M2000M GPU is the hardware available to train the model. This GPU is considerably slower than the NVIDIA RTX 2070 Super GPU, which is used by my colleague. For this reason, collaboration on the training of the model on the NVIDIA RTX 2070 Super GPU might be necessary.

The programming of the model will be done using Python with the libraries Tensorflow and Keras [5], since these are most applicable for image machine learning processes.

To develop the program for the model to run on a Raspberry Pi 3 with monocular camera module, OpenCV will be used. OpenCV [5] is an open source computer vision and machine learning library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products.

5.3 Prototype

The prototype will consist of 3D printed glasses with an integrated monocular camera. The integrated camera is connected to the Raspberry Pi 3 powered by a portable power supply. The monocular camera will supply live video to the predictor to predict the gender of the person in front of the camera. An earpiece device will be utilized to inform the user about the gender.

Chapter 6

Conditions

The following conditions should be taken into account before and during the research:

SBC The SBC (single board computer) available for the prototype is a Raspberry Pi 3 with interated monocular camera, this limits the design possibilities for the prototype.

GPU The NVIDIA Quadro M2000M GPU is available for training. Compared to the NVIDIA RTX 2070 Super GPU the NVIDIA Quadro M2000M GPU has low computational power, resulting in longer training times.

Chapter 7

Planning

The expected time needed for the execution of each task is displayed in Table 7.1. Table 7.2 displays the back-up planning in case of retakes in week 21.

Week	Schedule
17	Elaboration on what a neural network is, which one to select and investigation of machine learning libraries
18	Selecting a data reduction algorithm that reduces the size of the required data set for training
19	Selection and Trial testing of machine learning algorithms for gender recognition
20	Selection and Trial testing of machine learning algorithms for gender recognition
21	understand Raspberry Pi module and set-up
22	24 Development of program for depth mapping
23	24 Development of program for depth mapping
24	24 Development of program for depth mapping
25	Trial testing and integration with Raspberry Pi
26	Design of the prototype
27	Drawing conclusions, report writing and improvements
28	Report writing and improvements

Table 7.1: Planning

Week	Schedule
17	Elaboration on what a neural network is, which one to select and investigation of machine learning libraries
18	Selecting a data reduction algorithm that reduces the size of the required data set for training and trial testing of machine learning algorithms for gender recognition
19	Selection and Trial testing of machine learning algorithms for gender recognition
20	Study for retakes
21	Retake FEM and Production Management
22	understand Raspberry Pi module and set-up
23	24 Development of program for depth mapping
24	24 Development of program for depth mapping
25	Trial testing and integration with Raspberry Pi
26	Design of the prototype
27	Drawing conclusions, report writing and improvements
28	Report writing and improvements

Table 7.2: Back-up planning in case of retakes

Week	Schedule
16	Getting familiar with hardware/software setup
17	Study for retakes
18	Study for retakes
19	Elaboration on what a neural network is, which one to select and investigation of machine learning libraries
20	Selecting a data reduction algorithm that reduces the size of the required data set for training and trial testing of machine learning algorithms for gender recognition
21	Selection and Trial testing of machine learning algorithms for gender recognition Retake FEM and Production Management
22	understand Raspberry Pi module and set-up
23	24 Development of program for depth mapping
24	24 Development of program for depth mapping
25	Trial testing and integration with Raspberry Pi
26	Design of the prototype
27	Drawing conclusions, report writing and improvements
28	Report writing and improvements

Table 7.3: Second back-up planning in case of retakes

Bibliography

[1] J. Andreu-Perez, F. Deligianni, D. Ravi, and G.-Z. Yang, “Artificial intelligence and robotics,” 06 2017.

[2] T. Davenport and R. Kalakota, “The potential for artificial intelligence in healthcare,” *Future Hospital Journal*, vol. 6, pp. 94–98, 06 2019.

[3] P. Ackland, S. Resnikoff, and R. Bourne, “World blindness and visual impairment: despite many successes, the problem is growing,” *Community Eye Health*, vol. 30, pp. 71–73, 01 2017.

[4] “Seeing ai.” [Online]. Available: https://www.microsoft.com/en-us/ai/seeing-ai

[5] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow*. O’Reilly Media, Inc., 2019.

[6] T. Soyata and N. Powers, *Face Recognition: A Tutorial on Computational Aspects*, 01 2016, pp. 405–425.

[7] P. Useche Murillo, R. Moreno, and J. Pinzón Arenas, “Comparison between cnn and haar classifiers for surgical instrumentation classification,” *Contemporary Engineering Sciences*, vol. 10, pp. 1351–1363, 12 2017.

[8] S. S. Liew, M. Khalil-Hani, F. Radzi, and R. Bakhteri, “Gender classification: A convolutional neural network approach,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 24, pp. 1248–1264, 03 2016.

[9] X. Dong, S. Chen, and S. Pan, “Learning to prune deep neural networks via layer-wise optimal brain surgeon,” 05 2017.