Facial recognition from RGB-Depth images (identification)

Authors
Ilmar Uduste, xudust00@fit.vut.cz
Eslem Güler, xguler03@stud.fit.vutbr.cz
Christian Johansen, xjohan04@stud.fit.vutbr.cz
GitHub Repository

Abstract

Facial recognition and identification is a popular topic in today's world. Most of the methodologies behind recognising people in pictures rely on RGB images, but it is also possible to take it up a notch and use depth sensors for a more accurate representation and classification of the human face. In this project, an Intel Realsense Depth Camera is used to capture RGB and depth images simultaneously and a convolutional neural network is built to classify the identity of the person in the picture.

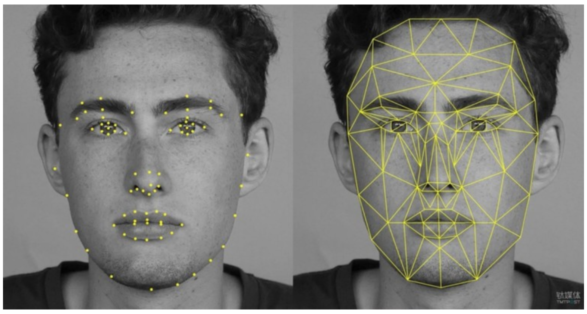

Figure 1: An illustrative example of facial recognition and feature extraction.

1 Introduction

Many applications nowadays rely on computer vision for facial recognition and identification. The most common use of this could be unlocking your phone by looking at it (e.g. FaceID), but it's also used in surveillance technologies and the fight against crime. Sometimes, though, computer vision can be wrong and identify the wrong person and the consequences of this could be severe (citation needed). To circumvent this issue, depth information is used in addition to usual approach of using RGB images in facial identification for a more accurate measure of facial features. This depth information needs to be captured at the same time and from the same angle as RGB data and to achieve this, depth sensors are usually placed right next to regular cameras.

2 Task Definition

The aim of this project was to collect RGB-D data and build a facial recognition and identification application that could **a)** recognise a face from an input photo (Figure 1) and **b)** identify the person in the photo. The people possibly recognised by the application will have to have their pictures in the training data of the model as well, so the "identification scope" of this application is just three people (the three people who made this project and consented to have their pictures taken and used for training a facial identification model).

In collecting the RGB-D images, an Intel Realsense Depth Camera had to be connected to a computer with a development environment. Also, a new convolutional neural network had to be trained in order to identify people from RGB-D image pairs. Lastly, the performance of the identification application was evaluated to see if it is better than a random guesser which, in this case, would be correct 33% of the time.

3 Methodology

The RGB-D images that were used for as the basis for training and testing data were collected via an Intel Realsense Depth Camera. A plethora of issue was had with connecting the camera to MacOS systems, but it finally worked while using Parallels to run Ubuntu and connecting the camera to a MacBook using a USB 2.0 cable (USB 3.0 didn't work). The depth image values each pixel with the distance in meters for that particular pixel. Bear in mind that these values are integers, so in a profile picture, the entire face might be just one contour. The inaccuracy of the depth image leads to the depth information being unusable in actual facial recognition or identification (Figure 2).

That leaves RGB images to be the only data that

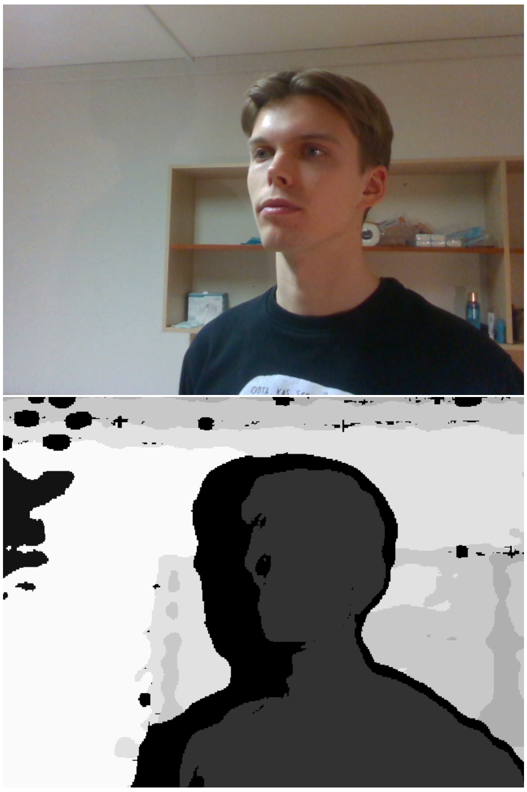

Figure 2: An example of an RGB and (normalized) depth image pair recorded by the Intel Realsense Depth Camera.

the facial recognition/identification model can be trained on. Approximately 60 photos were taken of each author from various angles and lighting conditions while the author was displaying a range of emotions. The initial plan was for the application to also detect the person's emotions, as this has previously been done (Szwoch and Pieniążek (2015); Ijjina and Mohan (2014)), but this was deemed too difficult and data-intensive for our project in the end. Each author's RGB-D image pairs were saved to a folder with their name on it and this data can be downloaded in the project repository using the `download_sample_pictures.sh` script.

In order to standardise the input into the facial identification convolutional neural network, the RGB images were first scanned with a facial recognition model (dlib's frontal face detector), then the images were cropped accordingly to fit only the faces of the persons in them and finally, the cropped images were also scaled to a standard size of 170x170 pixels (Figure 3) by the `crop_faces.py` script.

And finally, the cropped faces act as both training and test data to the convolutional neural network that attempts to identify a person. The training and test data split has been chosen to be 80/20 in light of the low amount of data available for this project.

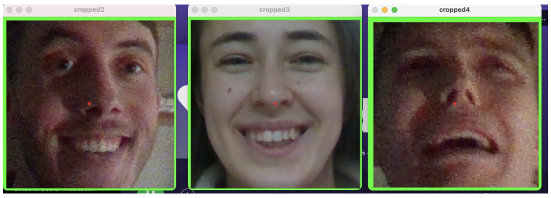

Figure 3: Images of the faces of the three project authors cropped and scaled to the same size.

4 Results and Analysis

Facial recognition worked 100% of the time on the RGB images, as this was done using a pre-established and proven model of dlib's frontal face detector. The results of the facial identification model are unknown, as the implementation of it is not yet ready, therefore we cannot make any conclusions as to the accuracy and feasibility of facial identification using RGB images.

Although it is possible to say that using the Intel Realsense Depth Camera is not suitable for tasks which require accurate depth readings, as it gave depth information with an accuracy of a meter, which definitely is not accurate enough for use in facial recognition or identification.

5 Conclusion

Facial recognition is a well established task in computer vision, but taking it a step further to identification requires more work to pull off. In this project, it was showed that using depth information in addition to RGB images is possible, but not while using the Intel Realsense Depth Camera, as its depth accuracy is not enough to have any meaningful info around the area of the face. In its case, other depth sensors should be used (e.g. Kinect) to get more accurate depth readings along with RGB images of people's faces to use them in facial recognition and identification tasks.

References

Earnest Paul Ijjina and C. Krishna Mohan. 2014. Facial expression recognition using kinect depth sensor and convolutional neural networks. pages 392–396.

Mariusz Szwoch and Paweł Pieniążek. 2015. Facial emotion recognition using depth data. pages 271–277.