Transfer Learning Deep Convolutional Neural Network for RGB-D Face Recognition

by

Puveneswari D/O Shunmugam
(1930613121)

A thesis submitted in fulfillment of the requirements for the degree of Master of Science in Mechatronic Engineering

Faculty of Electrical Engineering Technology
UNIVERSITI MALAYSIA PERLIS

2020

ACKNOWLEDGEMENT

To God be the glory for how far he has brought me, giving me the strength and wisdom to carry out this research. I would like to express my profound gratitude to my supervisor, DR. KAMARULZAMAN BIN KAMARUDIN, for his understanding and professionalism in guiding me throughout this work. Very special thanks go to my family, especially my parents who gave me words of encouragement, support, and prayers. Lastly, to all my friends and those who have been involved directly or indirectly, thank you very much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
DCNN	Deep Convolutional Neural Network
EER	Equal Error Rate
FMR	False Match Rate
GPU	Graphical Processing Unit
IMU	Inertial Measurement Unit
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LBP	Local Binary Patter
PIFR	Posture Invariant Face Recognition
QP	Depth Local Quantized Pattern
RGB-D	Red, Blue, Green and Depth
ResNet	Residual neural networks
SDK	Software Development Kit
SIFT	Scale-Invariant Feature Transform
SVM	Support Vector Machine
VGG	Visual Geometry Group

LIST OF SYMBOLS

$f_o(x, y)$	Original image the
$f_b(x, y)$	Blurred image
$f_s(x, y)$	Sharpened image
R, G, B	Pixel value before normalization
r, g, b	Pixel value of the point after normalization

Pemindahan Pembelajaran Rangkaian Neural Pelingkaran Dalam untuk Pengecaman Wajah Berasaskan RGB-D

ABSTRAK

Pengecaman wajah dua dimensi telah dikaji selama beberapa dekad yang lalu. Dengan perkembangan pemindahan pembelajaran dalam rangkaian baru-baru ini, pengecaman wajah dua dimensi telah mencapai tahap ketepatan pengiktirafan yang mengagumkan. Walau bagaimanapun, masih terdapat beberapa cabaran seperti variasi posisi, pencahayaan pemandangan, emosi wajah, kehadiran oklusi wajah yang wujud dalam pengenalan wajah dua dimensi. Masalah ini dapat diselesaikan dengan menambahkan Imej kedalaman sebagai input kerana memberikan maklumat berharga untuk membantu memodelkan batas wajah dan memahami ciri wajah dan memberikan corak frekuensi rendah. RGB-D lebih mantap berbanding Imej RGB sahaja. Malangnya, kekurangan set data muka RGB-D yang besar untuk melatih DCNN adalah sebab utama penyelidikan ini tidak diterokai lebih awal. Sekarang dengan pendekatan pemindahan pembelajaran, beberapa penyelidikan telah dilakukan baru-baru ini untuk pengecaman wajah RGB-D. Sebagai sumbangan pertama, penyelidikan ini juga membina set data wajah RGB-D yang baru di bawah pelbagai cabaran wajah (pencahayaan, oklusi, emosi, dan pose wajah) menggunakan Kamera Kedalaman Intel RealSense D435 yang mempunyai resolusi kedalaman yang lebih baik berbanding dengan Kamera Microsoft Kinect. Sebagai sumbangan kedua, kajian ini mengembangkan model Twin-DCNN berdasarkan model Inception-ResNet-V2 dan VGG16 yang mengambil Imej RGB-D sebagai input. Aliran RGB (Inception-ResNet-V2) memproses gambar RGB sementara aliran Kedalaman (VGG16) memproses gambar Kedalaman secara berasingan. Seluruh lapisan atas dalam setiap model yang dilatih sebelumnya telah dipertahankan dan lapisan bawah disempurnakan dengan menambahkan lapisan yang terhubung sepenuhnya ke setiap model. Kemudian kedua-dua model DCNN aliran RGB dan Kedalaman digabungkan bersama. Akhirnya, lapisan Soft-Max ditambahkan dengan 50 kelas output. Model Twin-DCNN yang dibangunkan mencapai ketepatan 96% pada pangkalan data RGB-D yang baru dibina.

Transfer Learning Deep Convolutional Neural Network for RGB-D Face Recognition

ABSTRACT

Two-dimensional face recognition has been researched for past few decades. With the recent development of Deep Convolutional Neural Network deep learning approaches, two-dimensional face recognition had achieved impressive recognition accuracy rate. However, there are still some challenges such as pose variation, scene illumination, facial emotions, facial occlusions exist in the two-dimensional face recognition. This problem can be solved by adding the Depth images as input as it provides valuable information to help model facial boundaries and understand the facial features and provide low frequency patterns. RGB-D images are more robust compared to only RGB images. Unfortunately, lack of large RGB-D face databases to train the DCNN is the main reason for this research to be unexplored earlier. Now with the transfer learning approach, few researches have been done very recently for RGB-D Face recognition. As the first contribution, this research constructed a new RGB-D face database under various face challenges (illumination, occlusion, emotion, and face poses) using the Intel RealSense D435 Depth Camera which has better Depth resolution compared to the Microsoft Kinect Camera. As the second contribution, this research developed a Twin-DCNN architecture based on Inception-ResNet-V2 model and VGG16 model which takes RGB-D images as input. The RGB stream (Inception-ResNet-V2) processes RGB images while the Depth stream (VGG16) process Depth images separately. The entire upper layer in each pre-trained model has been maintained and the lower layer were finetuned by adding a fully connected layer to each model. Then both RGB and Depth stream DCNN models were concatenated together. Finally, the Soft-Max layer was added with 50 output classes. Developed Twin-DCNN model achieved 96% accuracy on our newly constructed RGB-D database.

CHAPTER 1 : INTRODUCTION

1.1 Research Background

Face recognition is a system for recognizing an individual's face from a picture, where an algorithm would be set up to comprehend the primary features, or some other significant attribute of the face, to match new face pictures with existing ones. Throughout the years, face recognition has stayed a famous field of research in machine vision field and Artificial Intelligence (AI) field. Regardless of the critical measure of research directed in face recognition, face recognition under different illuminations, expressions, face poses and facial occlusions have stayed both a hypothetical and viable problems to the date.

Recently, a Depth camera was developed by Intel and named as RealSense™ Depth Camera D435. Intel RealSense Technology is an item scope of Depth and following innovations intended to give device Depth observations capacities (Bock, 2018). The device was first acquainted with use in the 3D filter, self-ruling automatons, and robots. However, later on, research in the field of computer vision discovered applications for it since it has both a visual and infrared sensor. The Intel® RealSense™ D435 camera gives precise pixel by pixel shading and Depth of the object. This sensor discovery brought about numerous papers being studied in proving how the sensor can be utilized to make new computer vision algorithms and methods. The fact that Depth pictures can capture a good quality, 1280×720 pixels, provides the opportunity of utilizing Depth maps for 3D feature extraction (Bock, 2018).

Another field that has conquered the computer vision is Deep Convolutional Neural Networks (DCNN), which has fundamentally improved the cutting edge in AI in numerous machine vision applications. Perhaps the most compelling motivation prompting the recognition of such strategies is the accessibility of many pictures in present-day computer upheld by Graphics Processing Unit (GPU) handling capacities and the accessibility of vast databases of information that can be utilized for their training. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015) was a significant benefactor in giving this colossal measure of information for the general image classification task. Recently, researchers have made databases accessible for detection and verification of scene and object (Lin et al., 2014). This permitted the publication of different Deep Convolutional Neural Network which introduced new and exceptionally productive approaches to extract feature and classify objects.

Unfortunately, in Deep Convolutional Neural Network and RGB-D face recognition, there is a big issue to obtain the large databases as the existing databases for Deep Convolutional Neural Network stay restricted to web sources which are two-dimensional images only. One example is the latest face recognition technique published by Google (Schroff, Kalenichenko, & Philbin, 2015) was trained to utilize 200 million RGB pictures with 8,000,000 typical personalities. It is understood that building this huge database beyond the capacities of most global research bunches in the world (Parkhi, Vedaldi, & Zisserman, 2015).

In conclusion, the advances in both RGB-D sensor and Deep Convolutional Neural Network (DCNN) has significant potential for solving many research challenges

of face recognition, for example, face recognition under extreme illuminations, emotions, occlusions, ages and pose variations.

1.2 Problem Statement

RGB based face recognition has been researched for more than 30 years, and it has recently achieved significant improvement with Deep Convolutional Neural Network (DCNN) development. However, RGB face recognition in an unconstrained environment is still facing many challenges to get high accuracy rates. Illumination variant, presence of occlusions, different lighting levels and not forgetting variant facial expressions are some of the major problems that still exist in 2D face recognition.

These problems exist as the RGB data only contains three colour channels and lack of Depth information. With the combinations of Depth image which is RGB-D data, an algorithm would be able to solve those challenges that still exist in RGB face recognition. Although some researches had ventured into deep learning-based RGB-D face recognition in recent years, the approaches are still limited and can be improved.

Besides, most of the researchers had used Kinect based RGB-D face database which technically has a lower quality of Depth resolution (512 x 424 pixel) and contains lots of noise such as database created by (Hg et al., 2012). A better good quality Depth resolutions (1280 × 720 pixels) RGB-D database under various face challenges would be desirable to ensure that a reliable RGB-D based DCNN model could be developed.

1.3 Research Objectives

The objectives of this research are described below:

I. To create a new face database that contains RGB, Depth and Point Cloud data under pose variation, illumination variation, emotion variation and occlusion variation challenges.

II. To develop a new Twin-DCNN architecture using two different pre-trained DCNN models for RGB-D face recognition.

III. To compare the accuracy of the Twin-DCNN model with the existing DCNN based RGB-D face recognition approaches.

1.4 Research Scope

This research is bounded to the following scopes and limitations:

I. The implementation of a Deep Convolutional Neural Network (DCNN) is based on the transfer learning approach.

II. Pre-trained DCNN models such as ResNet50, ResNet-18, InceptionResNetV2, VGG19, VGG16, XceptionNet and AlexNet will be explored in this research.

III. The RGB-D database is constructed using Intel RealSense Depth Camera D435 to capture the RGB image, Depth image and Point Cloud data.

IV. The database is constructed under a controlled indoor environment, which is inside a closed room with no windows, no noises and with controlled lamp lights.

V. The database contains data of 50 individuals (25 males and 25 females). The age range is between 12-50 only, and subjects are from multi race background such as Indian, Chinese and Malay.

VI. This research focuses only on four face recognition challenges: scene illuminations, pose variations, face emotions and facial occlusions.

1.5 Research Contribution

Contribution of this research study are as listed:

I. This research has contributed a new database consisting of RGB, Depth, and Point Cloud data under pose variation, illumination variation, different facial expression, and variant occlusions for 50 individuals.

II. A new Twin-DCNN architecture using two different pre-trained DCNN models was developed to improve the accuracy of RGB-D face recognition.

1.6 Thesis Organisation

This thesis has presented the research about the transfer learning Deep Convolutional Neural Network for RGB-D face recognition. For the clarity of presentation, the thesis is organised into five chapters including the introduction chapter, and can be seen as follows:

Chapter two has presented the literature review of this research. First part of this chapter describes face recognition in general and the steps involved in the face recognitions. The subsequent part describes the previous face recognition researches based on a Deep Convolutional Neural Network. Following that, this chapter describes challenges that exist in the RGB face recognitions system and reviewed the types of the low-cost Depth sensor and then the review of the existing RGB-D databases in the face

recognitions field. The final part of this chapter is about the RGB-D image-based approaches used in face recognition.

Chapter three has presented the methodology of this research. This chapter had explained the technique that will help in advancing the state-of-the-art research in face recognition, such as the utilization of RGB and Depth pictures with Twin-Deep Convolutional Neural Network (Twin-DCNN) procedures for face recognition. The main objective is to develop the Twin-DCNN model that is capable of performing face recognition based on RGB-D pictures that were acquired with the Intel RealSense D435 camera. Since training a DCNN requires a huge database, transfer learning and fine-tuning were employed as a reasonable solution.

Chapter four has presented the results and the discussion of this research. Three types of experiments were carried out in order to test the constructed database and Twin-DCNN performance. The first experiment was carried out with RGB image while experiment two was carried out with Depth image. VGG19, VGG16, ResNet, InceptionResNetV2 and XceptionNet pre-trained DCNN models were implemented. The final experiment was carried on using RGB-D database. This experiment conducted on AlexNet, ResNet and Twin DCNN models. The results are discussed in chapter 4.

Chapter five has summarised the research presented within the thesis and draws overall conclusions. Possible future improvements and enhancements are also suggested.

CHAPTER 2 : LITERATURE REVIEW

2.1 Introduction

First part of this chapter describes face recognition in general and the steps involved in the face recognitions. The subsequent part describes the previous face recognition researches based on a Deep Convolutional Neural Network. Following that, this chapter describes challenges that exist in the RGB face recognitions system. The next describes the types of the low-cost Depth sensor and then the review of the existing RGB-D databases in the face recognitions field. The final section of this chapter is about the RGB-D image-based approaches used in face recognition.

2.2 Face Recognition System

Facial recognition is about recognizing an individual using a face image, sketch, or video. Facial recognition has been researched thoroughly for an extensive period of time. In fact, facial recognition is a very popular research topic in the Artificial Intelligence field. Face recognition can be used in many applications such as phone verification, attendance system, surveillance system, banking application, enforcement and not forgetting the security system. Face recognition consists of a few basic steps. Figure 2.1 have illustrated general framework of face recognition.