Faculty of Information and Communication Technology

FEATURE BASED FACE DETECTION FOR UNCONSTRAINED IMAGES

Tioh Keat Soon

Doctor of Philosophy

2021

FEATURE BASED FACE DETECTION FOR UNCONSTRAINED IMAGES

TIOH KEAT SOON

A thesis submitted
in fulfillment of the requirements for the degree of Doctor of Philosophy

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

DECLARATION

I declare that this thesis entitled “Feature Based Face Detection for Unconstrained Images” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ……

Name : ……

Date : ……

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature	:	.....
Supervisor Name	:	.....
Date	:	.....

DEDICATION

I dedicate this work to:

MY PARENTS

Thank you for your love.

MY BELOVED WIFE

Thank you for your patience and encouragement.

MY SIBLINGS

Thank you for your support.

MY SUPERVISORS

Thank you for your patience, motivation and knowledge sharing.

MY SPONSOR (MALAYSIA GOVERNMENT)

Thank you for the scholarship support

ABSTRACT

Face detection for unconstrained images often encounter issues like background variation, pose variation, facial expression, occlusion and noise. Face detection utilises two main methods; feature based and image based methods. The feature based method benefits from rotation independence, scale independence and quick execution time as compared to the image based method. Feature based method utilises skin colour, facial and blob features. Current research on feature based method often emphasises on Viola Jones (V-J) face detection and is only limited to the in-plane rotation of positive or negative forty-five degrees. However, the utilization of V-J face detection with the inclusion of noise is a challenge because the image of other objects will often be mistaken for faces thus resulting in false detections. This thesis focuses on pose variation and noise challenges of unconstrained images and will cover three techniques for V-J face detection for unconstrained images, namely the combination of V-J face detection with rotation enhancements, Bicubic interpolation and ratio Scale Invariant Feature Transform (SIFT). In this thesis, these three techniques play different roles in face detection. The first technique begins with the rotation of the image file at thirty degree steps until it reaches a total rotation of three hundred and sixty degrees. At each thirty degree step, V-J face detection is applied, which in turn covers more angles of a rotated face. The second technique, Bicubic interpolation, corrects distorted images. The third technique, ratio SIFT, is a proposed post-processing to eliminate false detection for unconstrained images. Robust feature detection in scaling and invariant rotation is utilised in the above techniques to aid in the detecting of faces in images. Different face detections have been recommended for the unconstrained grey images and unconstrained colour images respectively with in-plane rotations and some with multiple faces. The images utilised for testing and evaluation in this thesis originated from Carnegie Mellon University (CMU) unconstrained grey images with in-plane rotations and Face Detection Data Set and Benchmark (FDDB) unconstrained colour images with multiple faces datasets. Fifty CMU datasets with twelve rotations on each image and various permutations resulted in six hundred test pattern images have been performed. Furthermore, another six hundred test pattern images from FDDB were also evaluated. These images have been measured through correct detection rate, true positive and false positive. The results from these measurements indicate that the proposed feature based face detection technique, focused on the V-J face detection method, for unconstrained images has the ability to detect rotated faces with high detection accuracy which in turn reduces false detections. In conclusion, the proposed enhancements will improve the current V-J face detection technique and overcome future challenges for unconstrained images.

PENGESANAN WAJAH BERDASARKAN CIRI DALAM IMEJ YANG TIDAK TERKAWAL

ABSTRAK

Pengesanan wajah untuk imej tidak berkekangan kerap menghadapi isu seperti variasi latar belakang, variasi ekspresi wajah, hampas dan gangguan tidak dikehendaki. Pengesanan wajah menggunakan dua kaedah utama; kaedah berasaskan ciri dan kaedah berasaskan imej. Manfaat kaedah berasaskan ciri adalah kebebasan putaran, kebebasan berskala dan masa pelaksanaan yang cepat dibandingkan dengan kaedah berasaskan imej. Kaedah berasaskan ciri mengguna ciri-ciri warna kulit, ciri-ciri wajah dan ciri-ciri gumpalan. Penyelidikan semasa terhadap pengesanan ciri kebiasaannya menekankan pengesanan wajah Viola Jones (V-J) terhad kepada hanya positif atau negatif empat puluh lima darjah putaran dalam satah. Walaubagaimanapun, penggunaan pengesanan wajah V-J yang terdapat cabaran gangguan tidak dikehendaki dalam imej akan tersilap sebagai wajah menyebabkan pengesanan palsu. Tesis ini memberi tumpuan kepada cabaran variasi ekspresi wajah dan gangguan tidak dikehendaki dalam imej yang berkekangan dan melangkumi tiga teknik bagi pengesanan wajah V-J untuk imej yang berkekangan, iaitu kombinasi pengesanan wajah dengan peningkatan giliran, interpolasi Bicubic dan nisbah Scale Invariant Feature Transform (SIFT). Dalam tesis ini, tiga teknik ini memainkan peranan yang berbeza dalam pengesanan wajah. Teknik pertama bermula dengan memutar fail imej dengan setiap langkah tiga puluh darjah sehingga tiga ratus enam puluh darjah. Setiap langkah tiga puluh darjah, pengesanan wajah V-J diterapkan, seterusnya meliputi lebih banyak sudut muka berputar. Teknik kedua, interpolasi Bicubic mengeluarkan imej kabur. Teknik ketiga, nisbah SIFT, dicadangkan sebagai pasca pemprosesan untuk menghapuskan pengesanan palsu dalam imej tidak berkekangan. Pengesanan ciri yang mantap dalam teknik penskalaan dan putaran yang digunakan dalam teknik-teknik di atas akan membantu dalam mengesan wajah dalam imej. Pengesanan wajah yang berbeza telah disyorkan untuk gambar kelabu yang tidak terkawal dan gambar warna yang tidak terkawal masing-masing dengan putaran dalam satah dan beberapa dengan beberapa wajah. Gambar yang digunakan untuk pengujian dan penilaian dalam tesis ini berasal dari gambar kelabu Carnegie Mellon University (CMU) tanpa putaran dengan putaran dalam satah dan gambar warna tidak terkawal Face Detection Data Set and Benchmark (FDDB) dengan set data pelbagai wajah. Data lima puluh CMU dengan dua belas putaran pada setiap imej dan pelbagai permutasi menghasilkan enam ratus pola uji gambar telah dilakukan.Selain itu, enam ratus lagi corak ujian imej dari FDDB juga telah dinilai. Imej-imej ini telah diukur melalui kadar pengesanan yang betul, positif benar dan positif palsu. Keputusan dari pengukuran ini menunjukkan bahawa kaedah teknik pengesanan wajah berasaskan ciri yang dicadangkan, tertumpu kepada kaedah pengesanan wajah V-J, untuk imej tidak berkekangan mempunyai keupayaan untuk mengesan wajah berputar dengan ketepatan pengesanan yang tinggi, seterusnya mengurangkan pengesanan palsu. Kesimpulannya, penambahbaikan yang dicadangkan dapat memperbaiki kaedah semasa teknik pengesanan wajah V-J dan mengatasi cabaran yang dihadapi untuk imej tidak berkekangan.

ACKNOWLEDGEMENTS

I would like express my sincere gratitude to my supervisors Professor Dr. Abd. Samad bin Hasan Basari for his patience, motivation, knowledge sharing during my research period. His guidance is always highly appreciated. I also would like express my sincere to my co-supervisor Professor Dr. Burairah bin Hussin for his valuable inputs.

Special thanks to my parents for their encouragements, my beloved wife for her patience and encouragements and my siblings for the support.

I also would like to thanks to all the staff in administration and lecturers at Universiti Teknikal Malaysia Melaka for their support.

Last but not least, I would like to thanks Malaysia government for the scholarship of MyBrain15 programme.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CMU	-	Carnegie Mellon University
DoG	-	Difference of Gaussian
DoH	-	Determinant of Hessian
FDDB	-	Face Detection Data Set and Benchmark
FP	-	False Positive
LBP	-	Local Binary Pattern
LoG	-	Laplacian of Gaussian
MSE	-	Mean Squared Error
NN	-	Neural Network
PNN	-	Probabilistic Neural Network
PSNR	-	Peak Signal to Noise Ratio
TP	-	True Positive
V-J	-	Viola Jones

LIST OF PUBLICATIONS

Soon, T.K., Basari, A.S.H., and Hussin, B., 2015. Image Duplication and Rotation Algorithms for Storage Utilization. *JTEC Journal of Telecommunication, Electronic and Computer Engineering*, 7 (1), pp.71–76.

Soon, T.K., Basari, A.S.H., and Hussin, B., 2015. Image Duplication and Rotation Detection Methods for Storage Utilization. *JATIT Journal of Theoretical and Applied Information Technology*, 79 (1), pp.1-6.

Soon, T.K., Basari, A.S.H., and Hussin, B., 2015. Image Duplication and Rotation Algorithms for Storage Utilization: A Review. *MEJSR Middle-East Journal of Scientific Research*, 23 (10), pp.2454–2458.

Soon, T.K., Basari, A.S.H., and Hussin, B., 2014. Enhancement of Rotated Face Detection and Image Duplication Methods. *MUCET Malaysia University Conference Engineering Technology*, pp.1-5.

CHAPTER 1

INTRODUCTION

1.1 Research background

Face detection is widely utilised in a multitude of preliminary applications especially to locate faces in a given image (Vezzetti and Marcolin, 2012; Naruniec, 2014; Lawrence and Bala, 2016; Marčetić and Ribarić, 2016). Face detection differs from other types of facial analysis algorithms like face recognition, facial expression recognition and head tracking (Nanni et al., 2014). Face detection is crucial and often being regarded as the first step of any facial analysis algorithm (Sanchez-Cuevas et al., 2013; Nanni et al., 2014; Pongakkasira and Bindemann, 2015; Dominic and Antony, 2016) utilised to identify an individual. Therefore, the accuracy of face detection is crucial because reliable face recognition is dependent on the basics of face detection (Mekami and Benabderrahmane, 2010; Naruniec, 2014).

Face detection methods have many modern-day applications. Recognition activity and tracking systems incorporate face detection methods to deal with airport security (Irgens et al., 2017). Modern cameras often come with built-in face detection methods to improve auto-exposure and auto-focus (Zafeiriou et al., 2015). In social media, Facebook utilises face detection mechanisms for face tagging (Zafeiriou et al., 2015; Witham, 2017). Face biometric modality are commonly integrated in mobile devices for face recognition and authentication (Vazquez-fernandez and Gonzalez-jimenez, 2016; Witham, 2017). Cetinkaya and Akcay (2015) mentioned that people counting systems are good security tools that enhance security and safety at educational facilities. Face detection can also be seen in

marketing methods to gather information on the number of customers that frequent a certain area (Hashemzadeh and Farajzadeh, 2016). Primary utilised in detecting the number of customers, businesses can strategically locate their products or maximise their advertisements' effectiveness by identifying locations that are most often frequented by people. Recently, face detection is developed to secure ATMs from physical and logical attacks. Face image is taken with sophisticated algorithms that analyse 4000 points of human face. The transaction service will be stopped if the ATM users are not detected. For instance, user is unable to complete withdrawal money without showing his or her face. It is advantage to bank to clarify who used the card for transaction in case of a dispute. Face tracking can be utilised by retailers to count the number of visitors and track the movement of visitors through their stores. The data can then be utilised to optimise store layout. Digital signage providers utilise face tracking to determine how many people viewing their displays. The data then be utilised to optimise their campaigns. The challenge is the video captured in a low frame rate. It might lose track of the face from one frame to another. The challenge can be overcome by setting minimum frame rate. Another good example of face detection is pose estimation. Estimating he head pose is beneficial in the automated guided cards, where an in-car device runs the head pose estimation algorithm to detect the drowsiness of the driver. Honda humanoid robot - ASIMO utilise motion and depth cue to reduce search space based on video based face detector.

Recently there are several commercial and non-commercial face detection products. For instance, DeepSight SDK is a face analysis based on deep learning models. DeepSight SDK provides the feature of face detection, face tracking and counting, multi-people analysis, age estimation, gender estimation, face blurring, attention time and data processed locally. InSight SDK provides analysing face videos in real time by offering seven emotional types such as neutral, surprised, sadness, anger, fear, happiness and disgust. Additional to

this, head pose, eye movements, gaze, age and gender. CrowdSight SDK can be utilised for face tracking. CrowdSight SDK can estimate the head pose, mood, age, clothing colour, gender and six general facial expressions. Face match usually utilised for access control, payment verification and cluster faces. Matlab and OpenCV are both computer vision library and can be utilised for face detection. Both libraries are based on V-J algorithm.

1.2 Face detection for unconstrained images

In most literature, unconstrained images are also known as uncontrolled environment or so-called in-the-wild image capture conditions (Orozco et al., 2015; Zafeiriou et al., 2015; Sagonas et al., 2016). Controlled environment capture conditions include specific parameters like restricted poses of a face, restricted face expressions, clean backgrounds, restricted lightings and being occlusion-free (Huang et al., 2007; Lenc and Král, 2015; Orozco et al., 2015). For instance, passport photos are taken in controlled environments (Haghighat et al., 2016) with predetermined lighting and background parameters. Face detection for unconstrained images remains very challenging, due to constraining factors (Marčetić and Ribarić, 2016). This thesis focuses on static images for unconstrained images.

There are many constraining factors (Marčetić and Ribarić, 2016) or variability of environment (Mekami and Benabderrahmane, 2010; Dominic and Antony, 2016) for unconstrained images that influence face detection. The below samples of unconstrained images from the CMU and FDDB dataset will demonstrate the constraints that influence face detection.

1.2.1 Background variation

Lenc and Král (2015) and Borovikov et al. (2018) define background variation. The colour of the face is similar to other objects in an image (Soria-Frisch et al., 2007). Figure

1.1 shows the background variations among the face, hills and river. Certain shapes, hills and river can be mistakenly detected as face region.

Figure 1.1: Background variation

Usually segmentation or thresholding will be utilised to remove the background colour to detect the face. However, in some circumstances, the background colour cannot be easily removed. Illumination is an example where lighting sources in particular can change the appearance of an image (Affonso et al., 2018).

1.2.2 Pose variation

Pose variation includes in-plane and out-of-plane face rotations (Burton and Bindemann, 2009; Zhang and Zhang, 2010) which are common during photoshoots. An example of these rotations are poses with various angles of headshots (Allaert et al., 2018). Figure 1.2 shows the pose variations.

Figure 1.2: Pose variation

Several people were lying down in different angles in one image and the faces are close to each other which results in difficult for face detection.

1.2.3 Facial expression

Allaert et al. (2018) define facial expression. Facial expressions can be defined as an individual's surprised, angry, fearful, awestruck and appalled look on their face or even just a smile. Closed and opened eyes also impair face detection when eye detection is utilised to determine a face. Figure 1.3 shows facial expression.

Figure 1.3: Facial expression

The smile that has exposed the teeth has impaired face detection. The face is toward negative angle from zero degree.

1.2.4 Noise

The characteristics of a camera like the lenses, sensors and temperature detector may result in the image being usually distorted (Zafeiriou et al., 2015; Borovikov et al., 2018). Figure 1.4 shows the noise in the original image.

Figure 1.4: Noise

The man put his face in his hands. The head and spectacle are toward negative angle. Objects can be mistaken for faces are caused by noise in the image, thus resulting in false detections.

1.2.5 Occlusion

Phimoltares et al. (2007) and Marčetić and Ribarić (2016) define occlusion. Sometimes, a face is partially covered by other objects like a facemask or hat. Certain face detection methods utilise all of the facial features on a face, this being the eyes, nose and mouth. Figure 1.5 shows the occlusion in an unconstrained image.

Figure 1.5: Occlusion

By partially covering the face, especially on eyes and most of its special features that make face detection more difficult.

Face detection basically utilises feature based methods or image based methods (Hjelmås and Low, 2001; Asteriadis et al., 2009; Hatem et al., 2015). Recent feature based method are Haar, skin colour, facial and blob features (Belaroussi and Milgram, 2012; Brancati et al., 2017; Prantl et al., 2017). The advantages of this method are its rotation independence, scale independence and quick execution time as compared to the image based method (Hu et al., 2011). This research aims to propose an enhancement of V-J face detection for unconstrained images based on Haar feature. They are rotation and noise.

1.3 Problem statements

Most often than not, human beings can easily detect a face in an image effectively compared to computers. Unless digital image processing systems have been properly programmed (Moallem et al., 2011). The image processing systems for a computer to detect faces in an image are much more complicated as compared to the naturally in-built recognition methods of human beings due to the nature of unconstrained images (Moallem et al., 2011; Aksasse et al., 2017).