SPHERICAL-BASED ITERATIVE CLOSEST POINT WITH NOSE TIP LOCALIZATION FOR UNCONSTRAINT VIEWPOINT 3D FACE REGISTRATION

BHUVENDHRAA RUDRUSAMY

UNIVERSITI SAINS MALAYSIA

2019

SPHERICAL-BASED ITERATIVE CLOSEST POINT WITH NOSE TIP LOCALIZATION FOR UNCONSTRAINT VIEWPOINT 3D FACE REGISTRATION

by

BHUVENDHRAA RUDRUSAMY

Thesis submitted in fulfilment of the requirements

for the degree of

Doctor of Philosophy

April 2019

DEDICATION

To my caring parents,
my lovely wife, and
my wonderful daughters.

ACKNOWLEDGEMENT

All praises go to Almighty God, either the darkest night or the brightest day, Your relentless blessing always felt in every second of my life, till the death takes place. Through You, countless people who directly or indirectly has contributed to the formation of this doctoral thesis, my sincere sense of gratitude to One and all.

I am grateful to have Associate Professor Ts. Dr. Shahrel Azmin Suandi as my supervisor, for nurturing me from a sketch paper to a compiled readable research dissertation. His guidance, patience, and faith resulted in the completion of this research; may this be the beginning.

Many thanks go to Ministry of Higher Education (MOHE) for providing financial support by granting MyPhD scholarship, indeed it is a relief.

Appreciation also goes to the member of my dissertation committee, Professor Ir. Dr. Nor Ashidi Mat Isa, Associate Professor Dr. Dzati Athiar Ramli and Associate Professor Iman Yi Liao who has generously given their time and expertise to make this work even better.

I am truly grateful to my research mates Dr. Abduljalil, Fauzi, Sami, Halina, Samsul, Faris, and Wasseem who has continuously scrutinized my work, and shared their memories and laughter at times. The common belief has definitely provided additional encouragement getting it done.

I must also acknowledge PHS1 librarians for entertaining my request over the years, enables me to reach the required reference books from various external libraries, I will be shorthanded without it. Also, for renting me a carrel room in the library where I found my solitude during the years of studies, it was my second home indeed.

Further thanks to my fellow friends, collogue, students, faculty members, and administrative staff who have shown friendship, given much hospitality, enlightened, entertained, knowledge and wisdom support over the years.

Last but not least, my wholehearted gratitude to my earnest family members for their unconditional love, encouragement, emotional support, understanding, and unbending faith; would not have achieved it without such a space.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

µm	Micrometer
2D	2-Dimensional
3D	3-Dimensional
3DMM	3D Morphable Model
AAM	Active Appearance Model
AICP	Affine Iterative Closest Point
ATM	Automated Teller Machine
CA-LapGF	Context-Aware Laplacian Regularized Gaussian Fields
CCD	Charge-Coupled Device
CCIF	Camera Image File Format
CMOS	Complementary Metal Oxide Semiconductor
CPD	Coherent Point Drift
DNA	Deoxyribonucleic Acid
DSMM	Direct Student's-t Mixture Model
EM	Expectation Maximization
ESICP	Efficient Sparse Iterative Closest Point
FOV	Field-of-View
GA	Genetic Algorithm
GB	Gigabyte
G-BSP	Generalized-Binary Search Partition
GMM	Gaussian Mixture Model
HCI	Human-Computer Interaction
ICNP	Iterative Closest Normal Point
ICP	Iterative Closest Point
ID	Identification Card

IRLS-ICP	Iterative Reweighted Least Square ICP
IT	Information Technology
KC	Kernel Correlation
LDA	Linear Discriminant Analysis
LieTrICP	Trimmed Iterative Closest Point Algorithm
LMM	Local Morphable Model
MAP	Maximum a Posterior
MINLP	Mixed-Integer Nonlinear Program
mm	Millimeter
NaN	Not-a-Number
PCA	Principle Component Analysis
PICP	Probability Iterative Closest Point
PSO	Particle Swarm Optimization
RAICP	Robust Affine Iterative Closest Point
RAM	Random Access Memory
RGB	Red Green Blue
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
RPM	Robust Point Matching
SA	Simulated Annealing
SFS	Shape-from-Shading
SGO	Stochastic Global Optimization
SICP	Spherical Iterative Closest Point
SICP+GA	SICP With GA
SICP-D	Denoised SICP
SP	Semidefinite Programming
STD	Standard Deviation

SVD	Singular Value Decomposition
TS-RPM	Thin Plate Spine Robust Point Matching
VFC	Vector Field Consensus
VR	Volume Rendering

TITIK TERDEKAT TERLELAR BERASASKAN SFERA DENGAN PENEMPATAN HUJUNG HIDUNG UNTUK PENDAFTARAN WAJAH 3D SUDUT PANDANGAN TANPA KEKANGAN

ABSTRAK

Pendaftaran wajah muka 3D merupakan suatu penyelidikan yang kritikal kerana ia boleh diguna pakai untuk mengabungan imej, pengenalan wajah muka, dan analisis gerakan. Ia merupakan suatu tugas yang mencabar untuk mendaftarkan imej muka sudut pandangan sisi disebabkan oleh oklusi sudut pandangan muka. Dalam keadaan sedemikian, kebiasaanya, imej wajah oklusi yang disebabkan oleh sudut pandangan ini dibaik pulih dengan menggunakan teknik imej pemprosesan sebelum pendaftaran muka 3D. Sebaliknya, pendekatan sedemikian adalah anggaran sahaja. Hasil sumbangan pertama kajian ini adalah titik terdekat yang berasaskan sfera (SICP) yang dicadangkan untuk pendaftaran wajah muka 3D sudut pandangan tidak terkawal sebelum sebarang bentuk pengubahsuaian atau pemulihan imej. Di samping itu, kaedah pendaftaran muka SICP dapat membantu penempatan hujung hidung dengan menggunakan algoritma genetik (GA). Kaedah SICP + GA yang dicadangkan dinilai berdasarkan pangkalan data wajah 3D umum yang merangkumi pelbagai pandangan wajah muka (termasuk sudut pandangan sisi sehingga kira-kira ± 90°). Kajian in dibandingkan dengan titik simpulan koheren (CPD) dan kaedah berasaskan titik paling dekat (ICP). Tinjauan hasil eksperimen menunjukkan bahawa SICP + GA lebih unggul berbanding dengan kaedah titik simpulan koheren (CPD) dan kaedah titik yang paling dekat (ICP), menghasilkan ralat pendaftaran min purata 1.0758 mm berbanding dengan 1.2465 dan 1.3540 mm, masing-masing. Lebih penting lagi, SICP + GA berjaya mendaftarkan kesemua imej pandangan depan dan sisi manakala CPD dan ICP gagal untuk mendaftar pandangan sisi apabila sudut mencecah kira-kira ± 90°. Di samping itu, penempatan hujung hidung kasar dengan bantuan kaedah pendaftaran muka SICP yang dicadangkan telah improvisasi jarak penempatan hujung hidung

dengan min purata 1.6767 mm dengan menggunakan GA. Selain itu, penemuan hasil eksperimen ini juga menunjukkan bahawa teknik yang dicadangkan amat sesuai apabila penempatan hujung hidung adalah tidak kepastian. Sebaliknya, imej wajah muka 3D agak hingar dan mengandungi data diluar lingkaran yang mencacatkan prestasi pendaftaran. Dalam usaha penyiasatan keberkesanan kaedah pendaftaran wajah yang dicadangkan terhadap kehingaran imej dan data diluar lingkaran ini, kaedah pasca dan pra-penapisan berasaskan teknik 2.5 σ yang sedia ada dan juga teknik kejiranan terdekatan yang dicadangkan digunapakai. Kedua-dua siasatan dijalankan secara berasingan dan hasil kajian menunjukkan bahawa kaedah pendaftaran muka SICP yang dicadangkan adalah mantap terhadap kehingaran imej and data diluar lingkaran. Juga diperhatikan bahawa terdapat banyak teknik ukuran yang digunapakai sebagai matriks untuk penumpuan kaedah pendaftaran wajah yang menyebabkan variasi dalam penilaian prestasi pendaftaran wajah muka. Atas dasar itu, teknik pengukuran prestasi yang sesuai telah dicadangkan agar menyelaraskan penilaian adil dan saksama bagi pelbagai kaedah pendaftaran wajah muka dicapai dengan mempertimbangkan teknik pengukuran jarak titik-ke-titik, titik-ke-baris, dan titik-ke-permukaan. Secara ringkasnya, kaedah pendaftaran wajah SICP yang dicadangkan telah menangani kelemahan wajah muka pendaftaran 3D pandangan sudut tidak terkawal.

SPHERICAL-BASED ITERATIVE CLOSEST POINT WITH NOSE TIP LOCALIZATION FOR UNCONSTRAINT VIEWPOINT 3D FACE REGISTRATION

ABSTRACT

3D face registration remains as a critical area of research as it can be applied for image fusion, face recognition, and motion analysis to name a few. It is a challenging task to register a viewpoint face image due to facial occlusion. Generally, the occluded face image due to viewpoint is enhanced and restored using generic image processing techniques prior to face registration. On the other hand, such an approach is approximate. The first contribution of this study is the proposed spherical-based iterative closest point (SICP) for unconstraint viewpoint 3D face registration prior to any kind of image enhancement or restoration. In addition, SICP face registration method able to assist nose tip localization using the genetic algorithm (GA). The proposed SICP+GA method was evaluated on publicly available 3D face database that provides various face poses (including viewpoint up to approximately ±90° yaw angle) and was compared against coherent point drift (CPD) and iterative closest point (ICP) based methods. The experimental finding suggests that the proposed SICP+GA is superior compared to coherent point drift (CPD) and iterative closest point (ICP) based methods, yielding an average mean registration error of 1.0758 mm compared to 1.2465 and 1.3540 mm of later two methods, respectively. More importantly, SICP+GA successfully registered all the frontal and non-frontal images while CPD and ICP failed to register profile posed when the viewpoint yaw angle was about ±90°. In addition, the proposed coarse-to-fine nose tip localization with the assistance of SICP face registration method using GA has improvised the nose tip localization distance with an average mean of 1.6767 mm. Also, the experimental finding shows that the proposed technique comes handy when the coarse nose tip annotation is approximate. Contrarily, the face images can be rather noisy and contain outliers which may skew the registration performance. In the effort of investigating the robustness of the proposed

SICP face registration method against noise and outliers, post- and pre-filtering methods were performed on SICP face registration method using existing $2.5\,\sigma$ and the proposed nearest-neighboring like technique, respectively. Both investigations were conducted independently and the finding from the statistical analysis suggests that the proposed SICP face registration method was robust against noise and outliers. Also, it was noted that there are many measurement techniques used as a matrix for the convergence of the face registration method that causes variations in performance evaluation. On that basis, an appropriate performance measurement technique was proposed that harmonizes fair evaluation for various face registration methods by considering major point-to-point, point-to-line, and point-to-surface distance measures. In summary, the proposed SICP face registration method has addressed the shortcomings for unconstraint viewpoint 3D face registration.

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Face biometric or commonly known as face recognition becoming popular traits among the biometric communities because it is a contact free, natural, psychological interest, and nonintrusive. Techniques applied to face recognition range from cooperative to non-cooperative method where the image can be taken from a controlled static mug-shot to uncontrolled face identification with a noisy background. Though 2-dimensional (2D) image-based face recognition system has been evolving for many years, it was noted still having challenges due to data variability such as illumination, occlusion, facial expression, and pose variation. Moreover, since the 2D image is based on the photometric measurement, the noises occurred in the face recognition system are rather sensor-dependent. On the other hand, the 3-dimensional (3D) image measurement is based on physical position.

The 3D-based face recognition has been major research in recent years primarily due to 3D data observation in pose variation and lighting conditions (Daoudi et al., 2013; Jourabloo & Liu, 2015). Also, partly, due to the advancement in 3D image acquisition system for face sensing such as laser-based active structured lighting, passive and assisted stereo imaging, and currently an infrared-based 3D image acquisition system (Beauvisage & Aouf, 2017). In addition, a reachable price tag has also driven the 3D biometric exploration even further (Zhang, 2012).

Numerous methods have been proposed in 3D face recognition and are generally categorized as Rapid Rejection of Candidates via Indexing, Matching via Alignment, Matching Deformable Models, Matching via Subspace Methods, and Matching via Local Surface Features (Flynn et al., 2008). A comparative study on 3D face recognition showed that

the Matching via Alignment and Matching via Subspace methods are far superior, which is Iterative Closest Point (ICP) and Linear Discriminant Analysis (LDA), respectively (Gokberk & Akarun, 2006). On that basis, the focus of this work is face alignment and related works are discussed as follows.

1.2 Problem Statement

There are three problem statements in this work as outlined in the sub-section below. The first problem statement requires two issues to be solved simultaneously as both are closely correlated in this work. The second and third problem statements complement this work.

1.2.1 3D Face Registration

Image registration works are given much importance as it can be applied for image fusion (integrate data from different measurement), object detection and recognition, object tracking, change detection, and motion analysis to name a few. By definition, image registration is the process of spatially aligning two images of a scene so that the corresponding points assume the same coordinate (Goshtasby, 2012b). In biometrics, the face alignment process is also known as face registration. An accurate face registration contributes to a significant performance to an overall face recognition as disclosed in (Sariyanidi et al., 2015).

Generally, face registration strategies can be categorized into point-based, part-based, and the whole face registration. The point-based registration localizes fiducial points which represent a shape of an object, for instance, a face, or even a feature expression and registers, respectively. The partial face registration (e.g mouth, eyes) localized each feature explicitly and registers appropriately. The whole face registration considers overall data points in the point cloud and computes the global geometrical transformation to align two faces accordingly. Face registration also can be divided into rigid and non-rigid based registration. The rigid-

based registration aligns two faces in the same coordinate system, while non-rigid based registration wraps the corresponding probe faces into reference faces accordingly. Take note that the word 'whole face rigid based face registration' used interchangeably as 'face registration' in this work.

Often problems arise when the images are taken from different sensors, photographs, viewpoint, depth, and time stamp. and the aim of face registration is to find the best corresponding point-pairs out of two images. Typically, the raw images are extensively preprocessed in removing the noises, smoothing the face surface, filling the surface hole, and restoring the missing parts. Also, the 3D data usually have holes or missing data in the concave area such as nostrils, eyes, and facial hair. One reason being, the 3D facial data is coarse (about 4000 points) compared to 2D data which has about 50 folds more (about 200,000 points) (Wang et al., 2007) and may not be suitable for image processing without considering the reference image. 3D face reconstruction also leads to similar issue if general restoration is done without fusing thee reference image (Jiang & Chen, 2008). Such issue would be worsened if both probe and reference images were heavily enhanced and may result with loss of important facial information (Maiseli et al., 2017) and put the individual into a general classification (Huang & Aizawa, 1993). Hence, Ansari and Abdel-Mottaleb (2005), Hernandez et al. (2015), and Jiang et al. (2018) recommends image enhancement on registered face.

Another noticeable challenges in face registration is viewpoint face registration. The common way of solving the viewpoint registration is to restore the missing information (occluded surface) using reconstruction algorithm. Three notable deficiencies in such approaches. (i) The use of generic face reconstruction algorithm may result in loss of facial details. Therefore, registering the probe face and fusing it with reference face would further enhance the face reconstruction (Hernandez et al., 2015; Jiang & Chen, 2008; Niswar et al., 2010; Nozawa et al., 2015). (ii) The face structure of a person is not symmetrical (Borod & Van Gelder, 1990). That said, reconstructing occluded faces due to viewpoint with the assumption of faces are symmetrical may result in poor facial restoration. (iii) Although it is

possible to reconstruct the missing surface of a profile pose due to viewpoint occlusion to some extent, such approach requires few pre-processing steps which may increase computation time for real-time face recognition.

Facial features such as eyes, eye corners, eyes brows, mouth corners, nose tip, nose bridge, and ears are vital source of information for face registration as it reveals intrinsic face information in form of mathematical model such as algebraic, geometric, statistical, differential, spatial, and spectral (Goshtasby, 2012a). Another key reason using the facial feature for face registration are geometrically stable and liable to noise. Typically, a face feature can be measured locally or globally and is denoted as facial keypoints. Some scholars used only one (Drira et al., 2013), three (Yu et al., 2016), 48 (Nair & Cavallaro, 2009), and up to 68 (Tulyakov et al., 2018) facial keypoints in biometrics works, respectively. Among these features, nose tip is highly sought for face registration. However, as viewpoint angle increases up to 90° yaw angle, surface-based nose tip localization method (Bagchi et al., 2012b) does not provide key information for nose tip localization. Moreover, Mian et al. (2007) and Song et al. (2014) reported that inconsistent nose tip localization across the faces also contributes to face registration error.

There are two major challenges discussed above; (i) viewpoint face registration, and (ii) nose tip localization. Both are closely correlated issues to be solved. Keeping these in mind, the purpose of this work is to register unconstraint viewpoint faces with nose tip localization without any kind of image enhancement to avoid any loses of face information.

1.2.2 Noise and Outliers

In actual practice, probe images are generally noisy and unobserved due to the viewpoint, uncontrolled illumination, and other occlusion. Hence, the probe images are largely enhanced to restore the critical information prior to face registration. Image enhancement methods such as filling the missing hole, remove the outliers, smoothing the face surface, and

etc. are generally involved. However, such approach is rather approximate since it is based on a general algorithm. Reason being, no specific ground truth was used as a reference for image enhancement. Hence, such approach may result in missing of critical information (Maiseli et al., 2017). Similarly, Hernandez et al. (2015) and Jiang et al. (2018) noted that a better face reconstruction can be performed on registered face as it is less bias to generic face model. On the other hand, the challenge is, the proposed unconstraint viewpoint face registration method should be able to perform on noisy conditions to mimic the real-world scenario rather than performing on an ideal face model. On that basis, another purpose of this work is to investigate the performance of the viewpoint face registration methods against noise and outliers.

1.2.3 Similarity and Dissimilarity Measure

Given two faces in same coordinate system, the similarity or dissimilarity measures the correlation and differences to quantify the outcome of a face registration, respectively. Many scholars proposed similarity or dissimilarity measure where each techniques has its own advantages and disadvantages in relation to discrimination powers and the sensitivity of the measure, not limited to the speed and accuracy of the proposed registration methods (Goshtasby, 2012c). The accuracy of registration error provides two information; (i) how close a probe face is matched to the reference face, and (ii) the difference between two faces. However, it may be bias if the results were to compare with different dissimilarity measurement for each method. Therefore, the third purpose of this work is to harmonize performance measure and fairly quantify the state-of-the-arts in this work.

1.3 Research Objectives

The aforementioned issues that are highlighted in this research motivate to propose face registration method that is able to handle viewpoint faces prior to any image enhancement or face restoration. Hence, this study has drawn following research objectives:

i. To propose a 3D face registration method with nose tip localization for unconstraint viewpoint including partially occluded faces prior to image enhancement or face reconstruction.

ii. To investigate the robustness of the proposed 3D face registration method against noise and outliers.

iii. To determine a fair measuring technique for the face registration methods.

1.4 Research Scope

Face registration is an interesting area of research and the research scope is confined to unconstraint viewpoint 3D face registration. No kind of image preprocessing, image enhancement or face reconstruction is considered in this work.

This work is also limited to the GavabDB (Moreno et al., 2003) 3D face database as it provides enormous viewpoint images (profile poses) approximately up to ±90° yaw angle. While some limited 3D profile face image database such as FRAV3D (Pardos) and York (Heseltine et al., 2008) is no longer publicly available due to personal data protection policy. Other publicly made available databases do provide neutral faces with various facial expressions and do not emphasize on viewpoint. Take note that the viewpoint is limited to maximum ±90° yaw and anything beyond, the image loses the facial keypoints and is not useful for face biometrics. In addition, this research does not consider external occlusion such as the subject wearing glasses or as such.

Although GavabDB face database provides real-world face images, the ground truth of the nose tip is not made available. It was noted that many works related to this database have opted to manual nose tip annotation due to an extreme viewpoint (Drira et al., 2013; Li et al., 2009; Mahoor & Abdel-Mottaleb, 2007; Yu et al., 2016). Hence, nose tip localization comparison is not presented in this work.

The research scope also confined within the rigid face registration, so the originality of the face structure is retained. Hence, the 3D face transformation in this work is limited to translation, and rotation only; while scaling is remained challenging for profile poses.

In this research, the proposed method is developed using MATLAB® R2014b software package with Intel® Core i5-3350 3.3GHz with 16GB RAM environment. Both hardware and software are provided by the learning institution.

1.5 Thesis Outline

This thesis is outlined in five chapters. The first chapter concisely presents the preliminary introduction of the 3D face registration with nose tip localization, effect of noise and outliers, and dissimilarity performance measure. The importance of the face registration is stressed and briefly discussed. The problem statements, research objectives, and research scopes are also offered. The remainder of this thesis is systematically presented as follows.

Chapter 2 presents a thorough literature review of this research. The related state-of-the-art techniques are compared and discussed immensely. The motivation and the rationale of this study are originated from the shortcomings of this literature review.

Chapter 3 presents the details of the proposed methods in this research. It begins with the proposed unconstraint viewpoint 3D face registration method. A coarse-to-fine nose tip localization using Genetic Algorithm (GA) is determined in this chapter. A detailed investigation on the robustness of the proposed face registration method against the noise and outliers is followed. An appropriate performance measurement technique is determined to fairly quantify the face registration error since both proposed and the competing methods have differences in the minimizing criterion. Lastly, a publicly available 3D image database for this work is determined.

Chapter 4 presents the experimental setup and the results of the proposed methods. The evaluation of the results is statistically quantified and is discussed thoroughly. In addition, the experimental is also qualitatively observed (visualized) to ensure the correctness of the face registration results across the competing methods. The significant of the propose method also statistically explored. All the experimental results are systematically analyzed and discussed.

Finally, Chapter 5 summarizes and highlights the contribution of this research with possible future works.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In the current world of globalization, the demand for security has increased and given much importance to prevent or protect from harm that causes vulnerable and/or valuable losses such as dwelling, person, community, organization or even to a nation. A security breach may cause a danger, threat, and risk results in high losses and therefore, having a security measure is a better option. For example, deploying a surveillance camera system may keep the intruders away from breaking the window in addition to protecting the valuable assets at home.

In general, biometric based authentication does not require a person to carry token or remember any additional information for verification such as username with secret password. The broad spectrum of unique human biometrics such as iris (retina), fingerprint, face, ear, hand geometry, hand vascular, gait, voice, palmprint, signature, or dental can be an alternative authentication (Jain & Ross, 2008). It is in line with rapid advancement in computer networking and transportation, communication field, and concerns highlighted on identity fraud and national security. Hence, there is a need for an efficient and reliable authentication system to ensure impostor does not gain privileges access associated with legitimately enrolled individuals.

One of the natural and most important human abilities used in their daily lives is face biometric and is known as face recognition among the research community.Various biometric modalities such as the face, hand, eye, finger, voice, and signature were studied for passport control system and the result shows the face biometrics has the highest compatibility (Heitmeyer, 2000). Woodard et al. (2006) carried out an experiment on 3D biometric modalities namely face, ear, and finger surface and suggested that face has the highest

identification rate. Hence, this motivates research in automated face recognition system especially in the uncooperative environment (Manjani et al., 2016), and is in line with human ability to recognize a person by simply looking at a person's face and able to identify quite accurately despite aging.

Compared to other biometric such as intrusive biometrics, face biometric gives an individual personal right and freedom (Jain, 2007) and hence, personal data is less exposed to security officers. The officer will only access the personal detail in case of false identification to further scrutinize and verify the data.

One such promising face recognition application is custom access at the airport. It can be automated using face recognition system as compared to manual verification where it can be flawed by the security officer. The officer in charge can request for further identification in case of identification mismatch. In such events, new data can be recorded for further improvement or be investigated if required. Nonetheless, with the research advancement in face recognition algorithm, such shortcomings can be developed progressively, be it 2D or 3D image-based content.

2.2 Face recognition

With the profound of human computer interface (HCI) system, face recognition has significantly coincides the lifestyle of modern society. Given the fact that machine is able to recognize the faces from still and video images, this research is spanning to numerous direction such as pattern recognition, image processing, computer vision, artificial intelligence, and etc. (Chellappa et al., 1995).

By definition, face recognition is a task that humans perform routinely and effortlessly in their daily lives (Li & Jain, 2011). The advantages of face recognition over other biometrics are easy to use, natural, and non-intrusive. The face biometric system is expected to operate

(i) face verification, and (ii) face identification in both still and video modes. Face verification is also known as face authentication. It is a one-to-one matching process where the query face image is compared with a template face for whose identity claimed. Face identification is also widely known as face recognition. It is a one-to-many matching process where the query face image was compared to face database to identify the closest match.

There are many advancements in the face recognition system since the introduction by Takeo (1973) and grows in-line with the progression of HCI system. Generally, face recognition can be successfully achieved under a controlled environment. Conversely, under an uncontrolled situation such as viewpoint, occlusion, illumination, expression, facial accessories and other such variables remain challenged. Such tasks become the main motivation and research progression for face recognition.

Traditionally, face recognition has been executed on 2D grayscale images and later followed by color-based images. Infrared images are also used due to high tolerance to illumination with addition hardware cost incurred (Hu et al., 2017). Along with the advancement of image acquisition system, many challenges was considered for real-time face recognition system (Garcia et al., 2017) and generally as follows (Martinez, 2009):

i. Face pose: Faces are in forms of complex 3D shapes and hence, large variation is expected when captured in 2D camera. Alternatively, a 3D camera can be deployed with a slight increase in cost and user cooperation much required at the moment.
ii. Facial expression: By all means, human are an emotional person and is expressed via large and complex movement of the facial muscles. Hence, facial expression results in distinct facial images.
iii. Illumination: A unique daily existence of nature cause different light ambient that causes distinct facial texture captured by the camera in the real world. In addition, different shape, pose, and sensor lighting cast with shadow causes further inconsistency in face texture.

iv. Occlusion: It is common to see people put on accessories on their faces such as glasses, facial make-ups, various hair style, hair color, mustache, beard, contact lenses, scarf, and etc. to match their clothing with attending event theme as part of their lifestyle. Hence, the face can be partially occluded and addition, it was noted that only 75% of viewpoint provides the most information.

The face recognition becomes more complicated when the list of challenges arises together. Nonetheless, it is one of the concerns the computer vision society is currently working on.

On the other hand, the advancement of processing power especially graphics hardware in personal computer and improvement over 3D image devices fascinates the interest of computer vision community in the recent years. Also, the 3D image acquisition system is now produced in a large quantity such as Kinect motion scanner for Xbox released by Microsoft and hence, the cost is reducing significantly. Therefore, the ecosystem for 3D face recognition system more or less is ready for the public to exploit and is becoming reality.

Based on a simple literature search on "2D face recognition" and "3D face recognition" using Scopus database (Tancock, 2018), it reveals that much interest noted in 3D face recognition among the research community and as shown in Figure 2.1. In addition, the trend is inclining in the past 10 years as compared to 2D face recognition which showed a

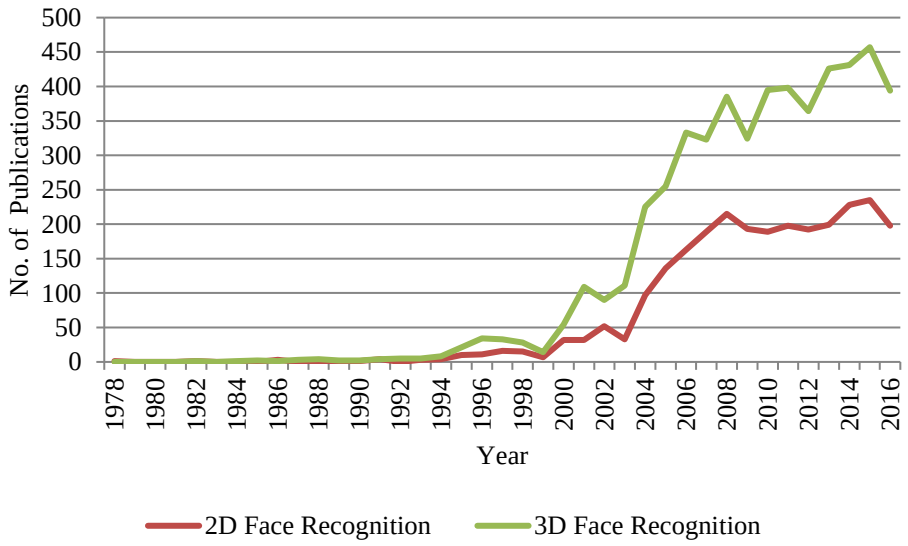

Figure 2.1. Trends in face recognition.

stagnant growth. Arguably the quality of 2D versus 3D works is not measured and left intentionally to be explored.

By definition, 3D face recognition is a method that exploits the 3D geometric information of the human face. 3D face recognition has better accuracy as compared to 2D by employing the features that are not sensitive to head orientation, skin texture such as makeup, differing facial expressions, and illumination (Kakadiaris et al., 2009). Like 2D face recognition, 3D face recognition can be categorized as face identification and authentication. In actual implementation, the common challenges of 3D face recognition are as follows:

i. Image database: Insufficient of 3D face image databases to verify the performances of 3D face recognition in various facial conditions and it should be publicly made available.
ii. Capturing devices: As most 3D devices are made for medical imaging, hence, it suffers from the depth, lower frame rate, artifacts, and the cost (expensive).
iii. Automation: It is expected the system is fully automated without any manual intervention such as a key landmark on the facial feature by the user.
iv. Robustness: Most of the images is captured under a controlled condition and should be robust enough to operate under various facial pose, expression, lighting, and etc.
v. Efficiency: Generally, 3D image generates a large amount of data. Hence, 3D data storage such as metadata is needed to be competitive with a 2D camera system.
vi. Accuracy: Accuracy gain of 3D over 2D face recognition system should be justifiable for commercialization.

A common 3D face recognition system is largely divided into two phases, (i) enrollment, and (ii) recognition. The enrollment steps are as follows:

i. Data acquisition: The Raw 3D image that was obtained from 3D vision sensor are generally preprocessed to improve device-specific issues.

ii. Alignment: The probe image is registered with respect to the reference image in the same coordinate system, be it either using the rigid or non-rigid method. This process is also known as image registration.

iii. Metadata: The data of registered probe image is reduced using feature or geometry map to obtain most significant coefficients matrix.

Figure 2.2 shows the major structure of the system in general (Li et al., 2013). The construction of a face recognition system may vary depending on specific applications. Take note that some systems may require feature extraction or feature point as a reference for image registration. Also, the term "image registration" is also known as "face registration" for face biometric tasks. Face recognition, either identification or authentication, can be directly achieved by comparing the probe metadata with the reference gallery using metric distance. Though 3D face recognition works sounds promising, there are many challenges yet to be answered. In this work, two key areas to improve 3D face recognition and is discussed in the following section.

Figure 2.2. General 3D face recognition system.

2.3 3D Face Registration

In general, the 3D image alignment process is known as point set registration. While in face recognition, the 3D face alignment process is known as 3D face registration and not to be confused with. By definition, the term registration is a spatially aligning process of two images to ensure corresponding points assume the same coordinate (Goshtasby, 2012b). The terminology "alignment of two images" manifests the geometrical transformation of image restoration such as identity, scaling, rotation, translation, horizontal and vertical shearing. It merely serves as dynamically merging the probe images into a static reference model. The

combination of these simple geometrical transformation techniques results in many image registration applications such as medical imaging (Audette et al., 2000; Fookes & Bennamoun, 2003; Savva et al., 2016; Shams et al., 2010), face recognition (Smeets et al., 2012), pose estimation (Ge & Fan, 2015; Meyer et al., 2015; Tulyakov et al., 2014), shape modeling (Danckaers et al., 2014), feature detection and many more.

Many has categorized image registration methods, Maintz and Viergever (1998) categorized it into eight categories as: registration basis, image dimensionality, geometrical transformation, optimization procedure, degree of interaction, subject, modalities, and object. The "registration basis" category refers to three basis of registration approach: (i) point-based methods, (ii) surface-based methods, and (iii) intensity-based methods. Registration basis is the top hierarchy of these eight categorizations. Following are short discussion of related categories for 3D face registration.

The "image dimensionality" category refers to number of geometrical dimensions of image space (Maintz & Viergever, 1998). In this word, 3D face registration refers to 3-dimensional image space. The 3D images are generally represented as volumetric data and is obtained using 3D data acquisition system. It is made of contiguous 2D images that provides a 3D array of image intensity values and is rendered into photo-realistic visualization. This dataset consists of voxels and is represented as points in a 3D space with certain position, usually points in a fixed grid. For simplicity of 3D image processing, these points are represented in x-, y-, and z-axis of Cartesian coordinates. Imagine x- and y- axis as 2D image plane, and z-axis as the depth image of intensity values, respectively. Hence, forms a volumetric data and these points is known as a point cloud.

The "degree of interaction" category refers to the control exerted by human operator over the registration algorithm (Maintz & Viergever, 1998). It may be a simple initialization or continuous assistance throughout the success of registration process. The "optimization procedure" category refers to continuous assessment of registration algorithm or function

(Maintz & Viergever, 1998), so it is minimized or maximized in closed-form fashion for global extremum solution, the termed also known as iterative search. The "modalities" category refers by which means the images to be registered (Maintz & Viergever, 1998). In other contexts, intramodal (monomodal) or intermodal (multimodal) registration.

More importantly, the "geometrical transformations" category refers to the mathematical form of geometrical mapping used to align the points in one space with those in the other. Let points from a space P of one view be mapped to the space Q. The transformation T applied on a point $p \in P$ transforms point p' as:

$$p' = T(p) \tag{2.1}$$

The point $q \in Q$ corresponds to p if the registration was equally successful. Otherwise, the displacement error $T(p) - q$ is known as registration error. Note that Equation 2.1 does not specify the behavior of transformation and is further categorized into two major categories as rigid and non-rigid transformations. The geometrical transformations categorization is based on the outcome of the image registration such that the straightness of lines is preserved or not, interrelated by means of "modalities".

Rigid transformation preserves the straightness of lines and angles between lines. Hence, the planarity of surface is preserved. It is simple registration that involves a translation and a rotation action. Using Equation 2.1, rigid transformation T defines a translation vector t and a rotation 3×3 orthogonal matrix R as:

$$p' = Rp + t \tag{2.2}$$

On the other hand, non-rigid transformation does not preserve either or both geometrical transformations qualities of a rigid transformations. Maintz and Viergever (1998) categorized the non-rigid transformation as follows: scaling, affine, projective, perspective, and curved transformation.

Scaling transformations is the simplest non-rigid transformation that preserves the straightness of lines and the angles. The transformations is given as:

$$p' = SRp + t \tag{2.3}$$

where $S = diag(s_x, s_y, s_z)$ is a diagonal matrix with scaling factor for each coordinate axes. It is also known as anisotropic scaling. In case of isotropic, $S = diag(s)$ where s is a positive scalar and $s = s_x = s_y = s_z$. The scaling results in enlarging and shrinking p and is known as similarity transformation.

Affine transformation preserves the straightness of lines and allow angles between lines to change. The transformations is given as:

$$p' = Ap + t \tag{2.4}$$

where there is no restriction imposed on elements of matrix A. Such transformation is appropriate when the image is skewed during acquisition and the correction is required.

Projective transformations also preserve the straightness of lines and allow angles between lines to change. The transformations is given as:

$$p' = (Ap + t)/(b \cdot p + \alpha) \tag{2.5}$$

A, t, b, and α can be folded into one homogenous 4×4 matrix such that:

$$u' = \begin{pmatrix} u_1' \\ u_2' \\ u_3' \\ u_4' \end{pmatrix} = Mu = \begin{pmatrix} a_{11} & a_{12} & a_{13} & t_1 \\ a_{21} & a_{22} & a_{23} & t_2 \\ a_{31} & a_{32} & a_{33} & t_3 \\ b_1 & b_2 & b_3 & \alpha \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ 1 \end{pmatrix} \tag{2.6}$$

The application of projective transformation is done by projecting 3D points to 2D plane and is known as perspective transformation (Hongwei et al., 2009).

Perspective transformations is subcategory of projective transformation which transform 3D points to 2D plane. Let $f = 1/|b|$ in Equation 2.5 and let $\hat{b}$ be a unit vector in the direction of the projective axis b, hence, obtained:

$$p' = fp/(p \cdot \hat{b} + \alpha f) \tag{2.7}$$

where $\alpha \neq 0$. Transforming p to a plane is achieved by zeroing the component of p' in the direction of b such that:

$$p' \rightarrow p' - (p' \cdot \hat{b})\hat{b} \tag{2.8}$$

This method is widely used for camera calibration (known as camera resectioning) to find the transformation of pinhole camera from 3D to 2D points in an image. The obvious application of image correction or calibration using perspective transformation are in medical imaging such as endoscopy, microscopy, laparoscopy, x-ray projection and other such image acquisition devices (Feldmar et al., 1995).

Curved transformations do not preserve the straightness of lines and allow angles between lines to change. The generic curved transformation is given as:

$$p' = \sum_{ijk}^{IJK} c_{ijk} x^i y^j z^k \tag{2.9}$$

where i, j, and k is the polynomial expression for x', y', and z' elements of p', respectively, and c_{ijk} is the coefficients. Generally, the polynomial expression I, J, and K is between 2 and 5, higher order polynomial cause spurious oscillation (Fitzpatrick et al., 2000). Methods such as splines, cubic splines, B-splines was derived for further improvement of general curved-based transformations.

To summarize on geometrical transformation, a rigid registration is applicable for an image that is sustained by shape during transformation. A non-rigid registration is applicable when the images are distorted during the acquisition process and have physical meanings of transformation, hence, it is rectified.

In general, since the 3D face images are not distorted, a rigid registration are considered for 3D face registration. Similarly, was considered in this work. Also, since such points are taken as reliable for image registration, point-based registration methods are considered and discussed as follows.

2.4 Point-based Registration

The point-based registration method is also known as point set registration. The ultimate goal of point set registration is to minimize the discrepancies between probe and reference image (point cloud) and hence, the probe image is geometrically aligned. The image registration formulation is broadly discussed, and Jian and Vemuri (2011) summarizes as follows. Let probe and reference image be finite-dimensional real vectors $\{P, Q\} \subset \mathbb{R}^n$, respectively. The objective of point set registration is to compute transformation T such that both $T(P, \theta)$ and Q are aligned, where θ is the parameter for registration algorithm. In general, the registration performance is measured in squared Euclidean distance and simplified as:

$$\underset{\theta}{\operatorname{argmin}} \left\{ \sum_{q \in Q}^{|Q|} \sum_{p \in P}^{|T(P,\theta)|} \|q - p\|_2^2 \right\} \tag{2.10}$$

where $|Q|$ and $|T(P, \theta)|$ are the cardinalities of Q and $T(P, \theta)$, respectively. The superscript and subscript indicates the squared Euclidean 2-norm of $\|q - p\|$. The fundamental problem of Equation (2.10) is the probe image contains noise, outliers, and missing data points and hence, is poorly performed. Maiseli et al. (2017) generalized the common challenges of image registration as follows:

i. alignment accuracy,

ii. outliers and missing data,

iii. dimensionality of datasets, and

iv. computational speed.

Since then, many techniques have been introduced to address the shortcomings as discussed in the following section and is limited to point-based registration methods only. Note that in this work, the word image registration is interchangeable to point-based registration and not to be confused with surface- or intensity-based registration.

2.4.1 Point Set Registration Methods

There is no one universal image registration technique fits all applications. It is limited by the geometrical characteristic of an image and depends on the local application problem statement. Two key elements in image registration are, (i) application, and (ii) fundamental technique. Many techniques proposed by the scholars and various aspects of image registration have been reviewed in the past (Li & Iyengar, 2014; Maiseli et al., 2017; Pottmann et al., 2006; Rodrigues et al., 2002; Salvi et al., 2007; Sariyanidi et al., 2015; Tam et al., 2013; Van Kaick et al., 2011; Zitová & Flusser, 2003). Maiseli et al., 2017 categorized point set registration methods in following discussion.

2.4.1(a) Iterative Closest Point Algorithm

One of the widely used image registration methods is iterative closest point (ICP) algorithm. The initial work were discussed in (Besl & McKay, 1992; Chen & Medioni, 1992; Zhang, 1994) and an overview structure of this algorithm is as shown in Figure 2.3.

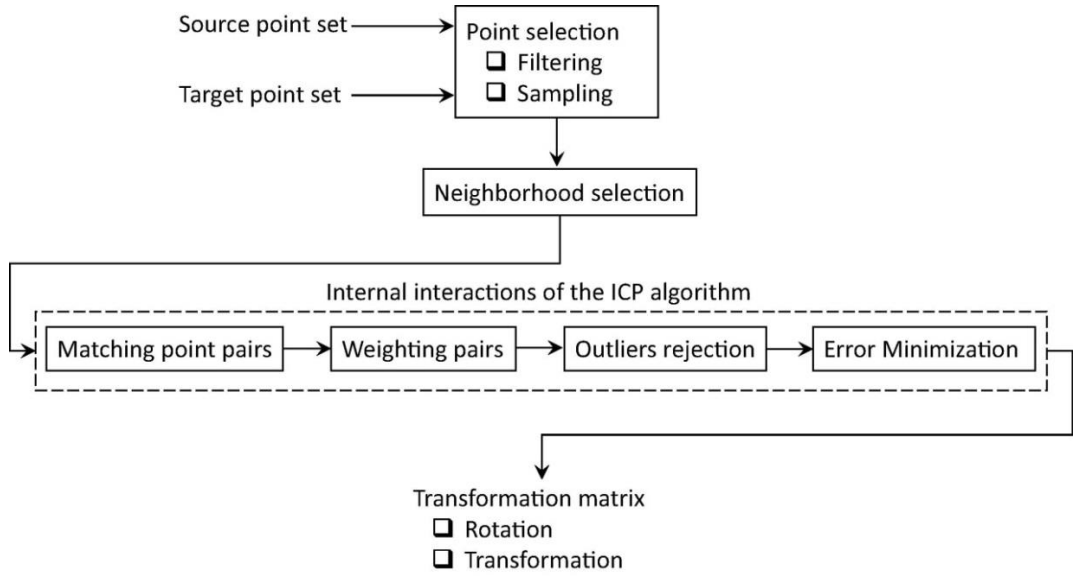

Figure 2.3. An overview of classical Iterative Closest Point (ICP) algorithm (Maiseli et al., 2017).

Generally, ICP is categorized as rigid registration method. Let the geometrical transformations of translation and rotation be denoted as T and R, respectively. Let probe and reference image be the source and target point cloud contained in finite-dimensional real vector as $\{P, Q\} \subset \mathbb{R}^n$. ICP is an iterative method that minimizes the mean squared error of the matching point pairs until it converged to optimum alignment such that P be registered with Q as following:

$$\underset{R,T}{\operatorname{argmin}} \left\{ \frac{1}{|Q|} \sum_{i} \|q_i - (Rp_i + T)\|_2^2 \right\} \tag{2.11}$$

where $q_i \in Q$ and $p_i \in P$ is the corresponding point pairs, $|Q|$ are the total number of data points in Q, and the superscript and subscript indicating the squared Euclidean 2-norm of $\left\| q_i - (R_{p_i} + T) \right\|$. The ICP algorithm is generalized to an arbitrary dimensional space and does not require any feature extraction. Many variants of ICP were proposed such as scaling-based (Chen et al., 2015; Du et al., 2010a), affine-based (Du et al., 2010b; Wu et al., 2016), sparse-based (Mavridis et al., 2015), Lie group parameterization (Dong et al., 2014), EM-based (Du et al., 2015), and point-based such as “normal point” (Mohammadzade &

Hatzinakos, 2013) to address some of the shortcomings in the applications. Rusinkiewicz and Levoy (2001) generalized ICP in the following tasks:

i. Points selection: The goal is to discard unwanted data point such as outlier if required.

ii. Matching: The goal is to select the probe and reference data point pairs for geometrical transformation.

iii. Weightage: The goal is to improve the performance of the ICP algorithm by introducing a controllable parameter.

iv. Rejection: The goal is to eliminate the point pairs that are likely to affect the minimization process.

v. Error minimization: To measure the convergence minimizing error.

There are many strategies adopted in "point selection". Besl and McKay (1992) used all available points, Turk and Levoy (1994) used uniform sampling, Masuda et al. (1996) used random sampling, Weik (1997) used gradient intensity value, and Rusinkiewicz and Levoy (2001) used distribution of normal, to name a few. The motivation of point selection strategy much depends on the application and outcome of the registration. Some of the application parameter to consider such as computation power, convergence rate, cost, robustness, distortion, precision, etc.

The "matching points" identifies which data point of the registering point cloud corresponds to that reference point cloud to be registered with, and it works hand-to-hand with "point selection" strategy. Besl and McKay (1992) proposed the matching of closest point in reference mesh, Chen and Medioni (1992) proposed the matching of intersection of the ray originating at the probe point cloud and of that normal point of the reference point cloud, Blais and Levine (1995) proposed the matching of source point cloud point with respect to camera (known as reverse camera calibration), Benjemaa and Schmitt (1997) proposed point-to-point

matching of probe to reference mesh, Weik (1997) proposed compatibility of intensity matching, to name a few. In most of the "matching point" techniques, some degree of exhaustive search is to be followed to ensure the corresponding point pairs results in optimal registration.

In addition, some ICP variant contains "weightage of pairs" parameter to provide some exception in matching process especially when the image is noisy and hence, least affected. Generally, the weightage value depends on the distance of the point-pairs. In such event, "rejecting pairs" was also introduced to eliminate unusual point-pairs with large distance apart and often seen in case of outliers. Such outliers can be identified for the point pairs with a standard deviation of more than 2.5 σ (Masuda et al., 1996). For example, Pulli (1999) proposed 10 % rejection rate. However, Rusinkiewicz and Levoy (2001) noted that although "rejection pairs" may have effect on stability and accuracy, it does not improve the convergence rate of image registration. On that basis, a constant weightage is provided, or both "weightage of pairs" and "rejection of pairs" parameters is ignored.

The "error metric" provides a measure for ICP minimization. The error metric can be in simple form of sum squared distances between corresponding point-pairs. Corresponding point-pairs could be in form of point-to-point, point-to-plane to name a few. To establish a closed-form solution, singular value decomposition (Arun et al., 1987), orthonormal matrices (Horn et al., 1988), quaternions (Horn, 1987), or dual quaternions (Walker et al., 1991) can be considered.

The ICP tasks from (i) to (v) are then repeated iteratively until one established a convergence rate where the minimization function becomes least meaningful. This is also to ensure the iteratively process ends in finite loop. Since then, many variants of ICP were proposed for various applications. The recent related ICP-based face registration works is discussed in later part of section 2.5.

2.4.1(b) Kernel Correlation Algorithm

Tsin and Kanade (2004) proposed kernel correlation (KC) algorithm for image registration. It uses correlation technique where the light intensity information of a probe and the reference image is analyzed to align the image appropriately. A set of entropy and affinity measure are estimated in formulating the correlation. The cost function of KC dynamically changes during the iteration and care required in selecting and formulating the kernel especially in the presence of outliers, occlusions, or clutters. To address the shortcomings, few variants of KC algorithm were proposed such as maximum kernel correlation (Tsin & Kanade, 2004), probabilistic estimation (Singh et al., 2004), robust surface matching (Fitch et al., 2005), multiple-linked similarity function (Chen, 2011), phase correlation fitting method (Tong et al., 2015) and multiple kernel (Nguyen & Wu, 2016). Moreover, the limitation such as dimensionality, transformation, kernel selection, and complexity are yet to be explored (Alam et al., 2010).

2.4.1(c) Gaussian Mixture Models Algorithm

The earlier works on Gaussian mixture model (GMM) with regards to image registration was found in KC algorithm (Tsin & Kanade, 2004) where GMM was intentionally used to avoid the stickiness of static data and to increase the robustness of the kernel. The fundamental implementation of GMM in image registration is to align two Gaussian mixtures and statistically optimize the discrepancies between them. Also, statistically is easily to interpret, manipulate, and formulate the framework appropriately. Since GMM has closed-form integral expression, the distance measure is computationally simple. However, this limits the usage in higher dimensions and remains challenging. Many scholars work around the GMM cost function to optimized image registration results (Arellano & Dahyot, 2012; Jian & Vemuri, 2011; Ju & Liu, 2012; Ma et al., 2016; Nguyen & Wu, 2013; Qu et al., 2017; Tao & Sun, 2015; Zhang et al., 2012; Zhang & Yang, 2016).