FORENSIC SKETCH TO MUGSHOT MATCHING ALGORITHM BASED ON DYNAMIC DIFFERENCE OF GAUSSIAN ORIENTED GRADIENT HISTOGRAM

SAMSUL SETUMIN

UNIVERSITI SAINS MALAYSIA

2019

FORENSIC SKETCH TO MUGSHOT MATCHING ALGORITHM BASED ON DYNAMIC DIFFERENCE OF GAUSSIAN ORIENTED GRADIENT HISTOGRAM

by

SAMSUL SETUMIN

**Thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy**

July 2019

ACKNOWLEDGEMENT

Alhamdulillah, all praise be to Allah, the Almighty, who granted me a good health, wisdom, and passion to complete this Ph.D research. I would like to record my appreciation to all those who directly or indirectly involved in this Ph.D work up to the stage where this thesis is completely written.

Foremost on the list, I would like to acknowledge and express my gratitude to my supervisor, Associate Professor Ts. Dr. Shahrel Azmin Suandi for his advise, encouragement, and exceptional guidance. I am indebted to him for always being with me while working towards my Ph.D. His vast knowledge in this research area had substantially helped me in realizing my ideas and even the ideas of him. His suggestions, insights and emotional motivations are really inspiring. Under his supervision, I have learned a lot. It is not too much if I say that Dr. Shahrel has definitely contributed to my intellectual development and drove my future as a whole. Besides my supervisor, I would like to thank the PhD proposal presentation panels: Associate Professor Dr. Haidi Ibrahim and Associate Professor Dr. Dzati Athiar Ramli for their encouragement, insightful comments, and hard questions. Next, my sincere appreciation also goes to Dr. Abduljalil Radman and Muhamad Faris Che Aminudin who have lent their hand in this research work.

I am particularly grateful for the assistance given by all academic, technical and administrative staffs in the School of Electrical and Electronic Engineering, Universiti Sains Malaysia (USM), who provide extraordinary support while this research is carried out. Furthermore, it is not possible to write every single name here but I insist to convey my special thanks to all my fellow postgraduate friends who consistently give me constructive suggestions and warm encouragement during my Ph.D study. My sincere appreciation extends to the Ministry of Higher Education and Universiti Teknologi MARA (UiTM) for the scholarship awarded to me under Skim Latihan Akademik Bumiputra (SLAB) as well as the Universiti Sains Malaysia (USM) for the fact that this research work was supported in part by the Universiti Sains Malaysia Research University Individual (RUI) Research Grant Scheme no. 1001/PELECT/814208 and 1001/PELECT/8014056, and the Universiti Teknologi MARA.

Not to forget, I would also like to express my gratitude to my beloved wife, Norsyira Zuraiza Omar, my loved kids, Syahmi and Syafiq, and all my family members for their endless emotional and spiritual support throughout my Ph.D research journey. It is hard to express the appreciation in words, but thank you for your understanding and always being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CCA	Cannonical Correlation Analysis
C-DoGOGH	Cascaded Static and Dynamic DoGOGH
CITE	Coupled Information-Theoric Encoding
CMC	Comulative Match Curve
CPU	Central Processing Unit
CUFS	CUHK Face Sketch Database
CUFSF	CUHK Face Sketch FERET Database
CUHK	Chinese University of Hong Kong
DAISY	Efficient Dense Descriptor
D-DoGOGH	Dynamic DoGOGH
DoG	Difference of Gaussian
DoGOGH	Difference of Gaussian Oriented Gradient Histogram
GB	Giga Byte
GW	Gabor Wavelet
HA	Horizontal Alignment
HAOG	Histogram of Averaged Oriented Gradient
HAVA	Horizontal and Vertical Alignment
HOG	Histogram of Oriented Gradient
IIIT-D	IIIT-Delhi Semi-Forensic Sketch Database

IoI	Identity of Interest
LBP	Local Binary Pattern
LDoGBP	Local Difference of Gaussian Binary Pattern
LLE	Local Linear Embedding
LRBP	Local Radon Binary Patern
MATLAB	MATrix LABoratory
MCCA	Multi-feature Cannonical Correlation Analysis
MLBP	Multiscale Local Binary Pattern
MRF	Markov Random Fields
MWF	Markov Weight Fields
OS	Operating System
PC	Personal Computer
PCA	Principle Component Analysis
PHOG	Pyramid Histogram of Oriented Gradient
PoI	Patch of Interest
PRIP	PRIP Hand-Drawn Composite (PRIP-HDC) dataset
RAM	Random Access Memory
S-DoGOGH	Static DoGOGH
SIFT	Scale-Invariant Feature Transform
SNS	Sparse Neighbor Selection
SRE	Sparse-Representation-based Enhancement

SSD	Spatial Sketch Denoising
SURF	Speeded Up Robust Features
SVR	Support Vector Regression
VA	Vertical Alignment

LIST OF SYMBOLS

a	patch index
$\vert . \vert$	magnitude
$\hat{a}$	neighboring patch index
α	number of bin
$*$	convolutional operator
b	bin
B	Gabor function
β	angle spaces
c	centre
C	oriented gradient histogram in cell
$\hat{C}$	scaling factor
χ^2	chi-squared
Σ	covariance matrix
d	distance
D	patch centroid target points
δ	selected distance
d_h	horizontal distance
d_p	maximum pixels coverage
d_v	vertical distance

dx	Haar wavelet response in horizontal direction
dy	Haar wavelet response in vertical direction
e	neighbor
E	number of neighbors
$\in$	element of
ε	small constant
exp	exponential
Eye_{CL}	center of the left eye
Eye_{CR}	center of the right eye
F	feature vector
f	frequency
f_a	feature vector of a patch
f'	normalized feature vector of a patch
F^B	GW feature vector
f^b	patch GW feature vector
F^H	HOG feature vector
f^h	patch HOG feature vector
f'^h	normalized patch HOG feature vector
$(.)^+$	operator to force negative values to zero
F^P	photo feature vector
F^S	sketch feature vector

g	gallery index
γ	sharpness of the Gaussian minor axis
g_s	sorted gallery indexes
G_σ	gaussian kernel
G_θ^σ	convolved oriented image gradients
G_x	gradient vector in x-direction
G_y	gradient vector in y-direction
H	image height in y-direction
I	image
I_{BW}	black and white image
$\hat{I}$	processed image
ι	counting index
J	neighboring points
K	number of photos in the gallery
k	number of short listed images
l	number of pixels off from the centre
L	number of resolution level
L_1	taxicab metric
λ	length of distance vector
M	number of patch
$\hat{M}$	number of target and neighboring patches

$Mouth_C$	center of the mouth
N	patch pixel length
P'	patch
P	photo
∂	partial derivative
ϕ	phase
$\hat{P}$	sampling point
π	pi
ψ	sharpness of the Gaussian major axis
Q	number of rings
R	radius
$\mathbb{R}$	real number
ρ	length of coordinate vector
s	scale
S	sketch
$\hat{S}$	complex sinusoid
σ	sigma
$\surd$	square root
$\sum$	summation
T	transpose
T_h	higher threshold

θ	angle
T_l	lower threshold
$\bigtriangledown$	gradient vector
U	convolution orientations
V	convolution scales
w	coefficient vector
W	image width in x-direction
$\hat{W}$	two dimensional Gaussian function
x	x-axis coordinate of pixel
y	y-axis coordinate of pixel
Z	feature vector length
ζ	white pixel coordinates

ALGORITMA PEMADANAN LAKARAN FORENSIK KEPADA GAMBAR TAHANAN BERDASARKAN HISTOGRAM PERBEZAAN GAUSSAN BERORIENTASIKAN KECERUNAN DINAMIK

ABSTRAK

Sistem dapatan semula gambar secara automatik berdasarkan lakaran wajah mempunyai aplikasi yang sangat berguna di dalam siasatan jenayah. Lakaran dan gambar wajah ini adalah dari modaliti yang berbeza. Dalam pendekatan pemadanan antara-modaliti, adalah tidak jelas penghurai yang mana paling modaliti-tak-berubah. Seterusnya, gambar dunia sebenar mungkin terdedah kepada variasi pencahayaan dan lakarannya mungkin mengalami sedikit tahap keterlaluan bentuk dengan butiran yang sangat kurang tepat. Dengan kesan ini, kadar dapatan semula berkurang dengan ketara. Dalam kerja penyelidikan ini, pada permulaannya, penghurai buatan tangan tempatan yang paling modaliti-tak-berubah ditentukan. Seterusnya, titik fidusial baharu untuk penjajaran muka dan penghurai buatan tangan tempatan baharu yang dinamakan Histogram Perbezaan Gaussan Berorientasikan Kecerunan (DoGOGH) diperkenalkan masing-masing untuk mengurangkan faktor keterlaluan bentuk dan untuk meminimumkan kesan pencahayaan. Ia diikuti dengan kaedah pengekstrakan ciri-ciri buatan tangan tempatan baharu yang dinamakan Histogram Perbezaan Gaussan Berorientasikan Kecerunan Dinamik (D-DoGOGH) dan Histogram Perbezaan Gaussan Berorientasikan Kecerunan Terlata (C-DoGOGH) untuk benar-benar mengambil kira kesan keterlaluan bentuk. Ketepatan dan kelajuan ditingkatkan lagi setelah menggabungkan lakuran ciri-ciri, Tampalan yang Berkepentingan (PoI) dan lakuran skor ke dalam kaedah yang dicadangkan. Keputusan kajian untuk Pangkalan Data Lakaran Wajah CUHK (CUFS) dan Pangkalan Data FERET Lakaran Wajah CUHK (CUFSF) menun-

jukkan bahawa kaedah yang dicadangkan mengatasi kaedah-kaedah yang terkini. Ia memberikan ketepatan pangkat-1 sebanyak 100% dan 95.48% masing-masing untuk pangkalan data CUFS dan CUFSF. Penilaian ini diperpanjang lagi kepada pangkalan data lakaran separa forensik dan forensik untuk menunjukkan bahawa kaedah yang dicadangkan tersaur untuk digunakan dalam penyiasatan jenayah dunia sebenar. Ia memberikan peningkatan ketepatan pangkat-1 sebanyak 28.56% dan 66.77% masing-masing untuk pangkalan data lakaran separa forensik dan forensik.

FORENSIC SKETCH TO MUGSHOT MATCHING ALGORITHM BASED ON DYNAMIC DIFFERENCE OF GAUSSIAN ORIENTED GRADIENT HISTOGRAM

ABSTRACT

An automatic photo retrieval system based on facial sketch has very useful application in criminal investigations. The face sketch and photograph are from different modality. In inter-modality matching approach, it is unclear which descriptor is the most modality-invariant. Next, the real-world photo may be exposed to lighting variation and the sketch may experience some degrees of shape exaggeration with very less accurate details. With these effects, the retrieval rate reduces significantly. In this research work, at the beginning, the most modality-invariant local hand crafted descriptor is determined. Next, a new fiducial points for face alignment and a new descriptor called Difference of Gaussian Oriented Gradient Histogram (DoGOGH) are introduced to reduce the factor of shape exaggeration and to minimize the illumination effects, respectively. It is followed by new feature extraction methods called Dynamic DoGOGH (D-DoGOGH) and Cascaded Static and Dynamic DoGOGH (C-DoGOGH) to really cater for the shape exaggeration effects. The accuracy and speed are improved further after incorporating feature fusion, Patch of Interest (PoI) and score fusion into the proposed method. The experimental results for CUHK Face Sketch Database (CUFS) and CUHK Face Sketch FERET Database (CUFSF) datasets demonstrate that the proposed method outperforms the state-of-the-art methods. It gives rank-1 accuracy of 100% and 95.48% for the CUFS and CUFSF datasets, respectively. The evaluation is extended further to semi-forensic and forensic sketch datasets to indicate that the proposed method is feasible to be used in the real-world criminal investiga-

tions. It gives rank-1 accuracy improvements of 28.56% and 66.77% for the semi-forensic and forensic sketch datasets, respectively.

CHAPTER 1

INTRODUCTION

1.1 Background

In law enforcement, traditionally, the process of searching potential suspects or the Identity of Interest (IoI) is performed manually. A large number of photographs need to be browsed by an eyewitness before selecting a few potential candidates. This process is very time-consuming and may not be accurate due to the fact that the environment may interfere with the eyewitness' focus, or they may experience fatigue while browsing the photographs (Shepherd, 1986). Apart from that, a forensic sketch is normally used as an evidence to find the suspect when there is no other evidence except the memory of the eyewitness. This sketch is rendered by a forensic artists based on the descriptions elicited from the eyewitness. Lois Gibson and Karen Taylor are well-known forensic artists involved in this kind of sketching (Sommer, 2015; Taylor, 2001). With the aid of eyewitness descriptions, the artists visualize the face in their mind and translate it into a pencil sketch by obeying a specific procedure as in Gibson (2008). This sketch is usually released to the public with the hope that the suspect can be identified based on the information retrieved from the public.

Assisting law enforcement to expedite the aforementioned process is one of the researchers' interests. This is done by matching the sketch at hand to photos in the mugshot database automatically. Successful matching method will allow for faster suspect apprehension. The identification process in a bigger picture can be seen in Figure 1.1. This research is ongoing since more than a decade ago and recently it

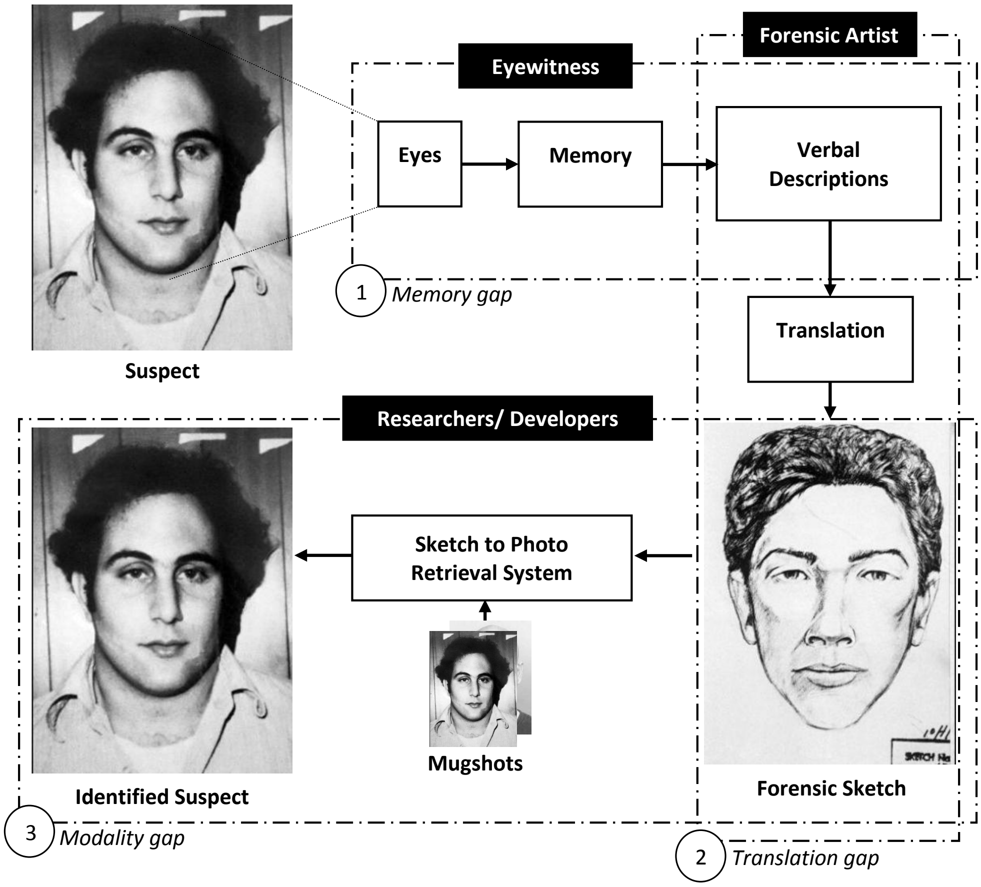


Figure 1.1: A bigger picture of the whole identification process (from suspect to identified suspect). All the involved stakeholders (i.e., eyewitness, forensic artist and researcher) are shown in the illustration. The images are from Klum et al. (2014).

has received much attention as the accuracy of existing methods remain significantly low. In order to match the sketch to photo automatically, the sketch is usually digitized using an electronic scanner before the matching algorithm can be employed. It is obvious that the sketch and photo are from different modalities. To perform image matching, different modalities introduce a substantial gap between the sketch and real photo (Tang & Wang, 2004). Matching facial sketches in different modalities is a very challenging task especially for automatic retrieval of face images.

Based on earlier researches, the gap can be minimized by performing a synthesizing process on either the sketch or photo to make up a pseudo-sketch/photo. When both images are in the same modality, feature extraction can be done with minimal gap. This is an intra-modality matching approach. Another approach used to match the faces is by extracting modality-invariant features on both images and the features are matched in that representation. This approach is called an inter-modality matching approach. Figure 1.2 shows the approaches. Among the two approaches, the intra-modality approach requires an advance synthesizing algorithm to generate synthetic images (i.e., pseudo-photo or pseudo-sketch). This is usually computationally complex. In addition to its complexity, although the transformation algorithms try their best to transform the image from one modality to another, the quality of the transformed images are still subjective as it also relies on the training samples used. Poor quality simply means lower matching accuracy. In fact, the synthesizing algorithms are often more complex than the recognition task itself (Han et al., 2013). In contrast, the inter-modality approach skips the above-mentioned complexity and focus on the actual recognition task. Based on that fact, this research work attempts to contribute in inter-modality matching approach. In this approach, for feature extraction, most of the researchers extract the feature using local hand crafted descriptor (Galoogahi & Sim, 2012a; Klare et al., 2011; Klare & Jain, 2013; Silva & Camara-Chavez, 2014). The local hand crafted descriptor refers to the descriptor that extracts manually designed features locally using the information present in the image itself.

From Figure 1.1, note that each process (i.e., the dotted line box) certainly introduces a gap. The gaps are memory (the loss of accurate details due to memory fidelity or the ability of a brain to remember the face accurately), translation (the misinterpretation of the verbal descriptions into the forensic sketch) and modality (the sketch and photo are generated using different medium). To address only on the modality gap,

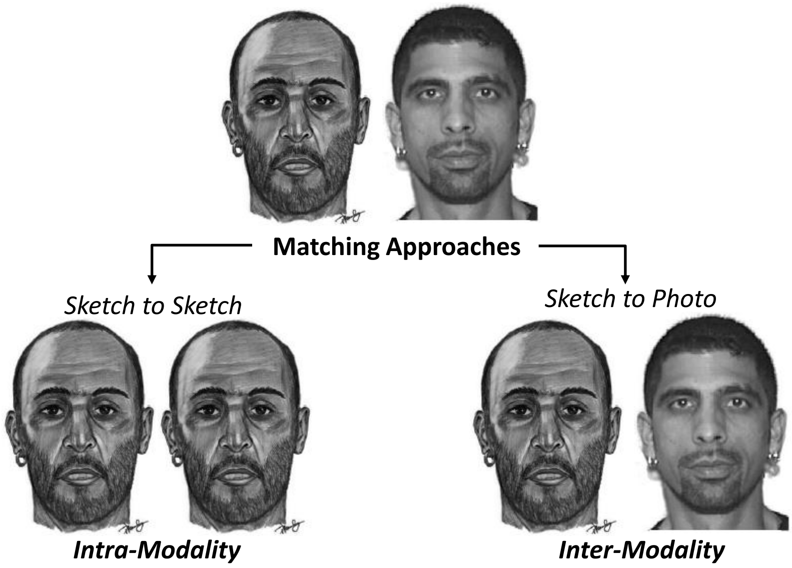


Figure 1.2: Two different matching approaches used in this research area. The images are from Klum et al. (2014).

viewed sketch is commonly considered by researchers. Next, to consider additional gap, it is extended to semi-forensic sketch such that the sketch is closer to forensic sketch. In summary, the facial sketches can be categorized into three categories that are viewed sketch, semi-forensic sketch and forensic sketch. All these sketches are studied to cater for a specific gap such that at the end the research objective and direction are toward real forensic application. Of all, viewed sketch to photo matching have shown a promising accuracy on clean dataset but still considerably low on dataset with illumination and shape exaggeration.

1.2 Problem Statements and Motivation

The main problem in matching sketch to photo is its accuracy due to the fact that the images are from different modality (Tang & Wang, 2004). To address this problem, researchers choose either a conversion process from one modality over the other (i.e., intra-modality) or using modality-invariant features (i.e., inter-modality) to match the images.

In inter-modality approach, seeking for modality-invariant feature becomes mandatory. It is because a correct feature representation may result in a high matching accuracy. To this end, global or local feature extraction method is commonly used. Out of these two techniques, local feature descriptors that extract the feature from patches have proven to be very effective in the sketch to photo matching (Alex et al., 2013; Galoogahi & Sim, 2012b; Klare & Jain, 2010; Klare et al., 2011; Klare & Jain, 2013; Klum et al., 2014; Silva & Camara-Chavez, 2014). Although, there are many local-based matching approaches have been proposed in the literature but the selection of normalization method, patch overlapping, patch size, distance measure, and feature descriptors are still ambiguous for inter-modality matching. Thus, it is crucial to identify a promising feature descriptor together with all the above-mentioned settings in this context.

Additionally, apart from having modality-invariant features to represent the images, the sketch and photo quality must also be taken into account because it may degrade the matching accuracy. To elaborate further, the sketch is drawn with no consideration of lighting conditions (i.e., no illumination) but it may suffer from slight shape exaggeration (especially on forensic sketch) (Zhang et al., 2011). While for photo, there is no possibility of shape exaggeration to occur but it has potential to be exposed to lighting variation. Disregard these imperfections will obviously sacrifice the performance.

Within the inter-modality approach, researchers mostly focus on seeking modality-invariant features to represent the image. In fact, shape exaggeration effects are generally neglected (Klare et al., 2011). Furthermore, the local features are usually extracted from static patches (i.e., the image is divided into some equal size of overlapping patches). Consequently, the extracted feature from a patch with exaggerated shape may not be accurate, and hence the similarity measure is made based on improper feature vectors. This may eventually degrade the matching accuracy.

Single feature representation seems to be insufficient to represent the images. Fusion of at least two features may result in stronger representation for better matching accuracy. In the field of multimodal recognition, feature fusion based on Cannonical Correlation Analysis (CCA) (Sun et al., 2005) has attracted researcher attention. The analysis attempts to find linear combinations of two different sets of feature vectors such that it maximizes the correlation between the two. This is done to increase the discriminative power. Based on this strategy, it becomes very popular and many CCA-based methods have been proposed in this research area (Correa et al., 2010; Haghighat et al., 2016; Li et al., 2015; Pong & Lam, 2014; Yang & Zhang, 2012). However, this approach does not really consider the effect of shape exaggeration and less minute details.

1.3 Research Objectives

The primary objective of this research is to propose algorithms which improve facial sketch to photo matching accuracy. The objectives include:

1. To determine a promising descriptor for feature extraction in the context of matching facial sketch to photo.

2. To introduce a new fiducial points for a better face alignment.

3. To develop a new feature descriptor to cater for illumination variations.

4. To propose a new feature extraction method which improves the matching accu-

racy by overcoming the shape exaggeration effects.

5. To improve the proposed new feature extraction method in terms of matching accuracy and processing speed.

1.4 Research Scope

The scope of this research work is limited to the development of algorithms for hand drawn facial sketch to mugshot matching under inter-modality framework. In general, the system comprises of image acquisition, pre-processing, feature extraction and classification stages. In this research work, the image acquisition stage is omitted because the benchmark datasets are ready for the researchers. Therefore, the research only focus on some pre-processing process, feature extraction algorithms and simple classification method. However, the main focus is at the feature extraction stage. For feature extraction, there are two main approaches to extract features: hand crafted feature (Galoogahi & Sim, 2012a; Klare et al., 2011; Klare & Jain, 2013; Silva & Camara-Chavez, 2014) and learned feature (Deng et al., 2018; Galea & Farrugia, 2017; Liu et al., 2018; Mittal et al., 2015; Parkhi et al., 2015; Wu et al., 2018; Zhang et al., 2015). In this research work, the contributions are within the hand crafted feature approach with untrained classifier. For the experimental evaluation, two widely used viewed sketch datasets are considered. The evaluation is also extended to two more datasets from semi-forensic sketch and forensic sketch categories. The proposed methods are evaluated on grey images because sketch is rendered without colour. In terms of the performance, Cumulative Match Curve (CMC) is used for retrieval rate evaluation. The proposed methods are expected to give high rank-1 accuracy from the CMC with reasonable processing time.

1.5 Research Contributions

The contributions of this research work are:

1. A promising descriptor in the context of matching facial sketch to photo is identified. The descriptor is better than the other evaluated descriptors.

2. A new fiducial points for face alignment are introduced. This minimizes the shape exaggeration effects.

3. A new descriptor for facial sketch to photo matching is developed, such that the illumination effects is addressed. The descriptor is called Difference of Gaussian Oriented Gradient Histogram (DoGOGH).

4. A new feature extraction method called Dynamic DoGOGH (D-DoGOGH) is implemented. The shape exaggeration effect is catered through this extraction method. To the best of our knowledge, no other local feature extraction method in the literature uses dynamic extraction. Then, a cascaded local feature extraction method called Cascaded Static and Dynamic DoGOGH (C-DoGOGH) is developed. It involves a static and dynamic feature extraction method. Employing C-DoGOGH significantly improves the matching speed.

5. Improving rank-1 matching accuracy by enhancing the CCA-based feature fusion matching method. This enhancement has been done by incorporating the DoGOGH, D-DoGOGH, and score-level fusion into the merely CCA-based feature fusion matching method. The matching processing time is improved by means of introducing the Patch of Interest (PoI) strategy, without sacrificing much on the accuracy.

1.6 Outline of the Thesis

There are a total of six chapters in this thesis. This chapter gives brief overview of sketch to photo retrieval system and draw attention to the important of this research. It is then followed by the problem statements, research objectives, research scope and research contributions. The rest of this thesis is organized as follows. In Chapter 2, the chapter begins with a concise explanation on the art of forensic sketch and the type of sketches used in this research work. This explanation is elaborated further to how those sketches are categorized. Then, a thorough review on the methods proposed in related publications is presented. Chapter 3 and Chapter 4 explain in details the proposed methods. Chapter 3 elaborates more on the newly proposed descriptor called Difference of Gaussian Oriented Gradient Histogram (DoGOGH) while Chapter 4 elaborates more on how the proposed descriptor is improved further to be Dynamic DoGOGH (D-DoGOGH), Cascaded Static and Dynamic DoGOGH (C-DoGOGH) and eventually be Cannonical Correlation Analysis (CCA) Fusion with D-DoGOGH on Patch of Interest (PoI). Chapter 5 discusses the experimental results obtained using the proposed methods on the benchmark datasets. Finally, a conclusion and future work of this study are drawn in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter begins with a concise explanation on the art of forensic sketch. The explanation is elaborated further to the type of sketches used in this research area. Then, a thorough review of existing methods in related publications is presented. Overall, this chapter is divided into ten sections. This section presents the overview of this chapter. In Section 2.2, the elaboration about sketch generation is presented. Subsequently, a brief explanation on the type of sketches is explained in Section 2.3. The previously proposed methods in intra-modality approaches and inter-modality approaches are discussed in Section 2.4 and Section 2.5, respectively. The literature is narrowed down to the feature extraction methods in Section 2.6 and followed by the local hand crafted feature extraction in Section 2.7. The matching process is discussed in Section 2.8. Section 2.9 reports the performance evaluation methods and eventually Section 2.10 summarizes the chapter.

2.2 The Art of Forensic Sketch

In face identification system, it is nearly impossible to identify suspect when there is no probe photo available (i.e., still or extracted frame from video) as a reference. However, there is an alternative solution to the aforementioned problem in which a forensic face sketch is created as the substitution. It is a special art drawn by a professional artist that is indeed differ from a normal face sketch drawn by a typical artist. This is because the drawing is solely drawn based on descriptions elicited from eyewitness' memory (Gibson, 2008). By having the forensic sketch, it will ease the identification process and eventually lead to suspect apprehension. In terms of the sketch generation, it initially begins with a sketch drawn by hand (i.e., hand-drawn) and followed by the use of computer application (i.e., software generated) to produce the sketch.

2.2.1 Hand-drawn

One of the techniques used to create a criminal face sketch is by sketching it on a paper using pencil. With the aid of eyewitness descriptions, the artists visualize it in their mind and translate it into a sketch by obeying a certain procedure as in the manual (Gibson, 2008). Lois Gibson and Karen Taylor are of the well-known forensic artists who involve in this kind of sketch (Sommer, 2015; Taylor, 2001). The sketch is eventually digitized using an electronic scanner for the purpose of face sketch to photo recognition. As for research advancement, face sketch datasets such as CUHK Face Sketch Database (CUFS) (Wang & Tang, 2009), CUHK Face Sketch FERET Database (CUFSF) (Zhang et al., 2011), and IIIT-Delhi Semi-Forensic Sketch Database (IIIT-D) (Bhatt et al., 2012) have been made available for public.

2.2.2 Software Generated

Another method to produce a forensic sketch is by generating it using a commercial facial composer (Han et al., 2013; Klum et al., 2013; Mittal et al., 2016; Ouyang et al., 2014). This is actually a computer software, which is used to render forensic sketches. With the aid of the eyewitness, a well-trained officer composes the face sketch by only selecting the shape or pattern options on screen. Although the sketches are far from perfect and having tendency to miss out important face details, it is an alternative technique to be considered. With computer technology advancement, this technique is anticipated to possibly replace the traditional forensic artists in the near future. E-FIT (E-FIT, 2013), EvoFit (Frowd et al., 2004), FACES (Faces, 2012) and IdentiKit (Identi-kit, 2012) are among the softwares that are used to generate the composite sketches of this kind. For the research use, publicly available dataset named PRIP has been prepared by Han et al. (2013) and later extended to e-PRIP by Mittal et al. (2014).

2.3 Facial Sketches

Retrieving a photo using its corresponding sketch is very challenging. It is even worse when the generated sketch has high degree of dissimilarity due to unavoidable constraint (e.g., memory gap, sketching styles and shape exaggeration) while generating the sketch. In this research, it is noticeable that researchers focus mostly on viewed sketch (refer Table 2.2) before handling real forensic sketch. This is due to the fact that viewed sketch is an ideal sketch in which there is no eyewitness gap interference exist while rendering the sketch. Hence, it is the most appropriate for baseline dataset. By analogy, perfect retrieval rate for viewed sketch should means the same for forensic sketch, but if not, it is due to the eyewitness gap. However, due to the substantial gap between the viewed and forensic sketch, a new type of sketch called semi-forensic sketch is introduced to bridge the gap (Bhatt et al., 2012). Not only that, recently, researchers have extended this boundary further by matching caricature sketch to its corresponding photo (Klare et al., 2012; Ouyang et al., 2014). The following subsections briefly explain each sketch type. Figure 2.1 shows where those sketches reside in face recognition domain while Figure 2.2 shows the example of the sketches. Note

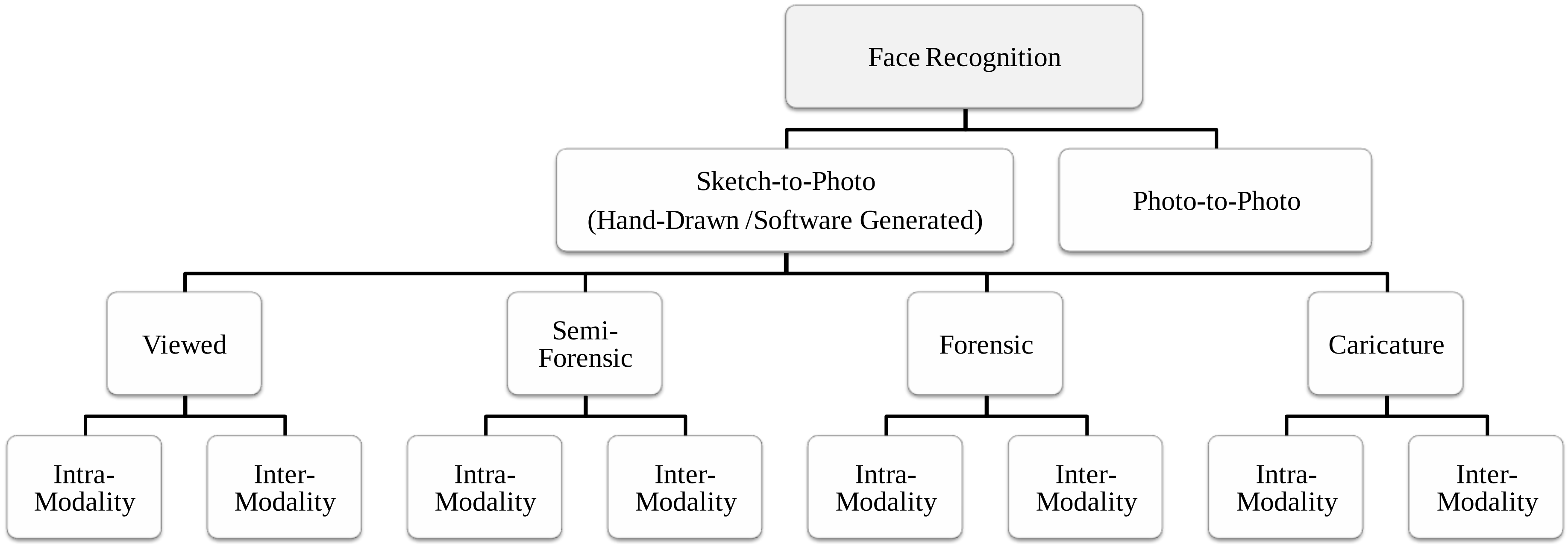
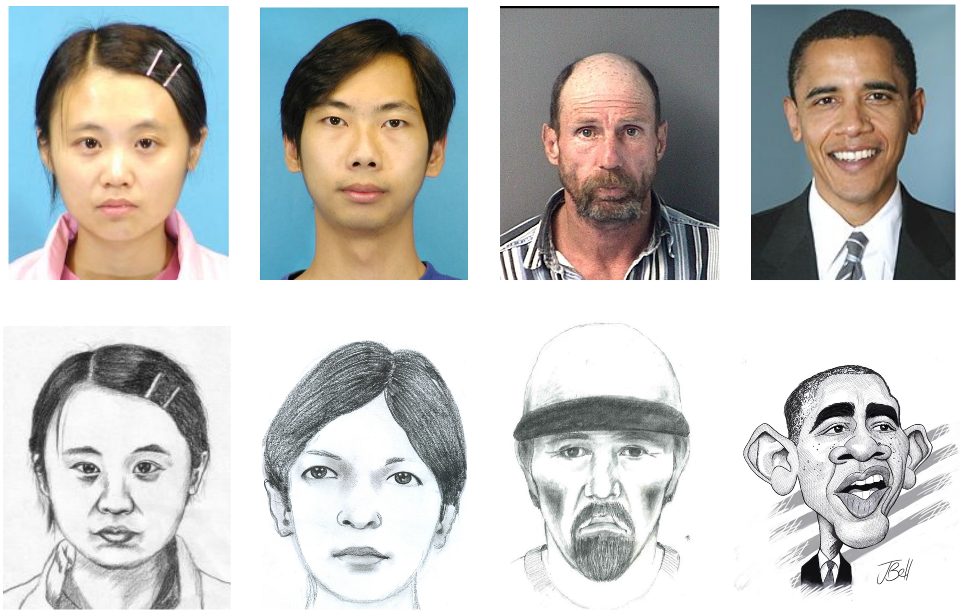


Figure 2.1: An overall overview of research contributions area for face sketch-to-photo matching.



(a) (b) (c) (d)

Figure 2.2: Examples of facial sketch pair; (a) viewed sketch (Tang & Wang, 2003; Wang & Tang, 2009), (b) semi-forensic sketch (Bhatt et al., 2012), (c) forensic sketch (Klum et al., 2014) and (d) caricature sketch.

that, in Figure 2.1, Photo-to-Photo subset is not within this research scope and the research mostly focuses more on the problems with regards to different pose, expression, illumination and occlusion (Tan et al., 2006).

2.3.1 Viewed Sketch

Sketching a face while viewing the photo or the subject is called viewed sketch (Gao, Zhong, Tao, & Li, 2008; Qingshan et al., 2005; Tang & Wang, 2004; Wang & Tang, 2009; Xiao et al., 2010; Zhang et al., 2010). CUFS (Wang & Tang, 2009) and CUFSF (Zhang et al., 2011) are the most popular dataset used in this category. Figure 2.2 (a) shows an example of viewed sketch pair.

2.3.2 Semi-Forensic Sketch

Face sketch drawn by an artist based on no descriptions from others but solely from their memory after observing the photo or the subject are defined as semi-forensic sketch. The third person is removed in this type of sketch to eliminate the effect of forgetting information. It has been prepared by Bhatt et al. (2012) and the dataset is named IIIT-Delhi Semi-Forensic Sketch Database (IIIT-D) (e.g., Figure 2.2 (b)). This is to bridge the gap between viewed and forensic sketches.

2.3.3 Forensic Sketch

In real world application, identifying suspect from a mugshot requires forensic sketch (as shown in Figure 2.2 (c), bottom) to be the probe image. The sketch is created based on verbal descriptions given by eyewitness (or the victim itself). It means the sketcher never see the suspect at all. From the descriptions, forensic artists visualize it in their mind and translate it into a sketch by obeying a certain procedure as in Gibson (2008), or a well-trained officer composes the face sketch by using a commercial composer software. Due to the gaps (i.e., memory and skill of the eyewitness and artist, respectively) while rendering the sketch, this type of sketch normally has a quality that is farther from perfect. Limited research works (Klare & Jain, 2010; Klare et al., 2011; Klare & Jain, 2013; Klare et al., 2014; Klum et al., 2013) have been done on this type of sketches due to limited dataset availability.

2.3.4 Caricature Sketch

A face sketch drawn with exaggerating the salient features or obvious characteristics of the subject is called caricature sketch. This type of sketch is said to be good if the viewer notice its likeness to the real subject even though it is not exactly alike. The example is shown in Figure 2.2 (d).

2.4 Intra-Modality Approaches

Generating a synthetic image in the preprocessing stage is categorized as intra-modality recognition technique. Intra-modality approaches uses synthetic sketch or synthetic photo as a replacement of original sketch or photo, respectively before state-of-the-art recognition algorithm is employed for identification. Most of the work proposed by Tang and Wang (Qingshan et al., 2005; Tang & Wang, 2004; Wang & Tang, 2009; Zhang et al., 2010) are under this category. It is then followed by Gao, Zhong, Tao, and Li (2008); Wang et al. (2011) and succeeding researchers (refer to Table 2.1 for details). The research works particularly focus on viewed sketch dataset. In terms of performance, the state-of-the-art has achieved 97.70% retrieval rate at the rank-1 as tested on CUFS dataset (Peng et al., 2017; Wang, Tao, et al., 2013). In order to get a clearer illustration, Table 2.1 tabulated the performance details for other existing methods. Next, for the convenience overview on intra-modality approaches, the methods are categorized into two main subsections; global approaches and local approaches (refer to Figure 2.3).

2.4.1 Global Approaches

Processing an image as a whole is referred to as global approach. As far as this approach is concerned, all proposed techniques in this category can be grouped into two subcategories; Eigentransformation and Direct Combined Model (as illustrated in Figure 2.3). Succeeding subsections discuss each of them.

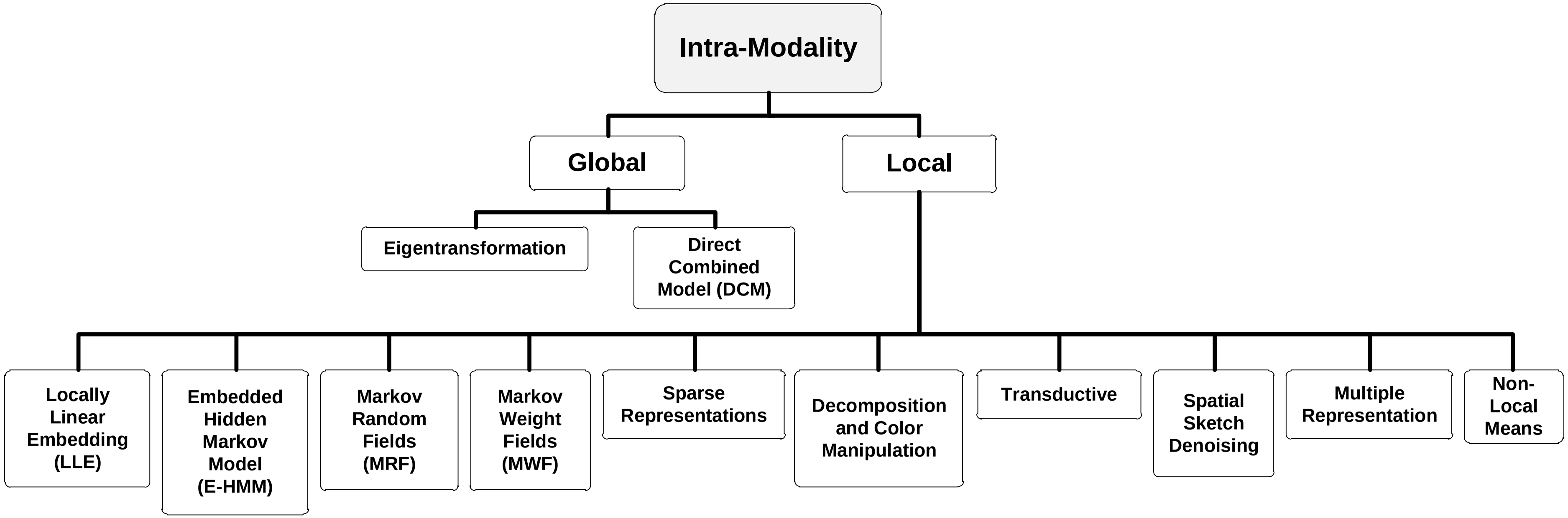


Figure 2.3: An overall overview for intra-modality approaches.

Table 2.1: Intra-Modality Approaches.

Authors (Year)	Database	Type of Sketch	[a]Hand/ [b]Soft. Gen	Synthesizing Technique	Local/ Global	Classifier/ Recognizer	Best Rank-1 (Rank-10) Accuracy (%)
Tang and Wang (2002, 2004)	CUHK	Viewed	Hand	Eigensketch transformation algorithm	Global	Eigenface (Turk & Pentland, 1991)	71.00 (96.00)
Tang and Wang (2003)	CUFS	Viewed	Hand	(Tang & Wang, 2002, 2004), but photo texture and shape are transformed separately	Global	Bayesian (Moghaddam et al., 2000)	81.30 (97.00)
Qingshan et al. (2005)	CUFS	Viewed	Hand	Local geometry preserving	Local	Kernel-based Non-linear Discriminant Analysis (KNDA) (Scholkopft & Mullert, 1999)	87.67 (99.00)
Gao, Zhong, Li, and Tian (2008); Gao, Zhong, Tao, and Li (2008)	CUHK	Viewed	Hand	Embedded Hidden Markov Model (E-HMM) and Selective Ensemble (SE)	Local	Eigenface (Kirby & Sirovich, 1990)	95.24 (-)
Wang and Tang (2009)	CUFS	Viewed	Hand	(Photo synthesis) using Multiscale Markov Random Fields (MRF) model	Local	Random Sampling LDA (RS-LDA) (Wang & Tang, 2004, 2006)	96.30 (99.70)

[a]Hand: Hand-Drawn ; [b]Soft. Gen: Software Generated ; - : Not Available or Not using CMC; $\approx$: Approximation; $*$: Refers to the corresponding database.

Table 2.1: Continued.

Authors (Year)	Database	Type of Sketch	[a]Hand/ [b]Soft. Gen	Synthesizing Technique	Local/ Global	Classifier/ Recognizer	Best Rank-1 (Rank-10) Accuracy (%)
Zhang et al. (2010)	CUHK, Celebrity	Viewed	Hand	Improved MRF from Wang and Tang (2009)	Local	Whitened PCA (Yang et al., 2005)	99.00 (100)
Xiao et al. (2010)	CUHK	Viewed	Hand	E-HMM (Gao, Zhong, Li, & Tian, 2008; Gao, Zhong, Tao, & Li, 2008) with image quilting	Local	Subspace learning methods (Kim et al., 2002; Pearson, 1901; Schölkopf et al., 1998; Sun et al., 2008) (Eigenfaces, KPCA, OTA)	100 (-)
Tu and Lien (2010)	Private	Viewed	Hand	Direct Combined Model (DCM)	Global	None	- (-)
Wang et al. (2011)	CUHK, *VIPSL	Viewed	Hand	Initial estimate: LLE (Qingshan et al., 2005). Enhancement: Multi-dictionary Sparse Representation	Local	Sparse Representation (Wright et al., 2009)	100 (-) *91.40 (-)
Zhang et al. (2011)	CUHK, *VIPSL	Viewed	Hand	Initial estimate: LLE (Qingshan et al., 2005), E-HMM (Gao, Zhong, Li, & Tian, 2008). Enhancement: SVR	Local	Sparse Representation (Wright et al., 2009)	100 (-) *94.00 (-)

[a]Hand: Hand-Drawn ; [b]Soft. Gen: Software Generated ; - : Not Available or Not using CMC; $\approx$: Approximation; $*$: Refers to the corresponding database.

Table 2.1: Continued.

Authors (Year)	Database	Type of Sketch	[a]Hand/ [b]Soft. Gen	Synthesizing Technique	Local/ Global	Classifier/ Recognizer	Best Rank-1 (Rank-10) Accuracy (%)
Gao et al. (2012)	CUHK, VIPSL	Viewed	Hand	SNS for neighbor selection and SRE for enhancement	Local	Sparse Representation algorithm	99.00 (-)
Liang et al. (2012)	CUHK	Viewed	Hand	BiCE Descriptor (Zitnick, 2010), MRF and coupled dictionary	Local	None	- (-)
Song et al. (2012)	CUFS	Viewed	Hand	Online coupled dictionary learning and global energy minimization, then MRF	Local	None	- (-)
Wang et al. (2012)	CUFS	Viewed	Hand	SCDL (learn a dictionary pair and a mapping function)	Local	None	- (-)
Li and Cao (n.d.)	CUHK	Viewed	Hand	Two-scale decomposition and color similarity map	Local	None	- (-)
Zhou et al. (2012)	CUHK, Celebrity	Viewed	Hand	Markov Weight Fields (MWF) and cascade decomposition method (CDM)	Local	PCA (Delac et al., 2005)	$\approx$ 89.00 ($\approx$ 99.00)

[a]Hand: Hand-Drawn ; [b]Soft. Gen: Software Generated ; - : Not Available or Not using CMC; $\approx$: Approximation; $*$: Refers to the corresponding database.

Table 2.1: Continued.

Authors (Year)	Database	Type of Sketch	[a]Hand/ [b]Soft. Gen	Synthesizing Technique	Local/ Global	Classifier/ Recognizer	Best Rank-1 (Rank-10) Accuracy (%)
Wang, Tao, et al. (2013)	CUFS, CUFSF	Viewed	Hand	Probabilistic graphic model and transductive learning	Local	Random Sampling LDA (RS-LDA) (Wang & Tang, 2004, 2006)	97.70 (99.70)
Wang, Li, et al. (2013)	CUHK, *VIPSL, visible-NIR	Viewed	Hand	Sparse Feature Selection (SFS) and Support Vector Reggresion (SVR) (Vapnik, 2013)	Local	Sparse Representation Classification (SRC) (Wright et al., 2009)	100 (-) *98.00 (-)
Song et al. (2014)	CUHK, AR	Viewed	Hand	SSD: Enhanced denoising algorithms (NLM (Buades et al., 2005) and BM3D (Dabov et al., 2007))	Local	PCA (Delac et al., 2005), Sparse Representation (Wright et al., 2009)	$\approx$ 89.00 ($\approx$ 99.00)
Peng et al. (2016)	CUHK, CUFSF, Celebrity, IIIT-D	Viewed, Forensic	Hand	MRF to learn multiple representations and alternating optimization strategy	Local	Random Sampling LDA (RS-LDA) (Wang & Tang, 2004, 2006)	97.70 (100)
Zhang et al. (2015)	CUFS	Viewed	Hand	Based on dictionary learning, sparse representation, greedy search strategy and MRF	Local	PCA (Delac et al., 2005)	$\approx$ 66.67 ($\approx$ 88.67)

[a]Hand: Hand-Drawn ; [b]Soft. Gen: Software Generated ; - : Not Available or Not using CMC; $\approx$: Approximation; $*$: Refers to the corresponding database.

Table 2.1: Continued.

Authors (Year)	Database	Type of Sketch	[a]Hand/ [b]Soft. Gen	Synthesizing Technique	Local/ Global	Classifier/ Recognizer	Best Rank-1 (Rank-10) Accuracy (%)
Chang et al. (2015)	CUHK	Viewed	Hand	NL-Means (Buades et al., 2005) coefficient and quilting (Efros & Freeman, 2001)	Local	None	- (-)
Galea, Christian and Farrugia (2015)	CUFSF	Viewed	Hand	*Intra and inter-modality fusion*. Combining Eigentransformation (Tang & Wang, 2004) with the Eigenpatches and then fused with Histogram of Averaged Orientation Gradients (HAOG) (Galoogahi & Sim, 2012b).	Local	PCA	44.05 (71.90)
Zhang et al. (2016)	CUFS, Celebrity	Viewed	Hand	Sparse representation-based greedy search (Zhang et al., 2015), multi-feature-based optimization model, cascaded regression strategy.	Local	Eigenface (Delac et al., 2005)	$\approx$ 58.67 ($\approx$ 84.33)
Chen et al. (2016)	CUHK	Viewed	Hand	Enhanced transformation algorithm as in Tang and Wang (2004)	Global	Enhanced Eigenface	72.00 (98.00)
Wang et al. (2017)	CUFS	Viewed	Hand	Combining both position constraint and global search. MRF	Local	Eigenface (Delac et al., 2005)	$\approx$ 72.67 ($\approx$ 94.00)

[a]Hand: Hand-Drawn ; [b]Soft. Gen: Software Generated ; - : Not Available or Not using CMC; $\approx$: Approximation; $*$: Refers to the corresponding database.

2.4.1(a) Eigentransformation

Tang and Wang (2002) and Tang and Wang (2004) have reported that the distance (i.e., for similarity measure) between a photo and sketch pair for the same identity is significantly larger than the distance between two photos of dissimilar identity. This has sparked to an outcome that direct application of state-of-the-art face classifier (i.e., eigenface - based on the article) to match different modality images is irrelevant. To tackle the aforementioned problem, the article proposed eigensketch transformation algorithm to convert a photo into a sketch so that both images are in the same domain. This is done before classification process.

In order to understand the proposed methods, let Q_s be a query sketch and M_p be the mugshot photos. Direct matching between Q_s and M_p will definitely give incorrect identity. Thus, eigensketch transformation is used to reconstruct a pseudo-photo, Q_p from the query sketch before matching procedure is executed (i.e., match Q_p and M_p). Based on a linear assumption as in Tang and Wang (2003), the pseudo-photo is generated by,

$$Q_p = \sum_{i=1}^{M} w_{s_i} T_{p_i} + \mu_p \tag{2.1}$$

where M is the number of photo in training samples, w_{si} and T_{P_i} are the contribution weight of each sample sketch image and the photo vectors training images, respectively. The mean μ_p is computed as $\mu_p = \frac{1}{M}\sum_{i=1}^{M} T_{p_i}$. In Equation (2.1), w_{si} is not from photo training samples, but it is obtained by projecting the normalized Q_s onto the eigensketch vectors of the sketch training samples. This is due to its linear approximation benefits. Once the Q_p is ready, in principle, the matching process is done by any state-of-the-art classifier. In addition to that, this algorithm is considered flexible because the comparison can also be made between sketch and pseudo-sketch (i.e., Q_s and M_s). To achieve this, instead of transforming the query sketch, Q_s into photo, transform all mugshot photos, M_p into a set of pseudo-sketch, M_s by using the same algorithm. Then, execute the matching procedure on the images. In terms of performance, the Comulative Match Curve (CMC) is used for evaluation. By computing this evaluation, eigensketch transformation approach shows better performance as compared to geometry (Wiskott et al., 1997) and conventional eigenface methods (Turk & Pentland, 1991). The best score is shown in Table 2.1. However, because of linear mapping between sketch and photo is being considered in this approach, improper face alignment may degrade the performance.

To avoid the problem and to further enhance the synthesis performance, Tang and Wang (2003) have extended the research work by employing separate transformation for both texture and shape information of the face photo. It is motivated by the fact that both texture and shape of a face pair are not identical. Also, fiducial mapping points between sketch and its corresponding photo is not that linear especially for complex facial structure. Thus, Tang and Wang (2003) demonstrates that the linearity can be improved by performing independent treatment to the texture and shape. The proposed algorithms extract shape information using face graph model and warp it to the mean face model (i.e., from the training set). By doing this, texture and shape are separated from the photo image. Then texture and shape for sketch are generated by exploiting eigentransformation technique. It is finally followed by execution of another warping process that manipulates the generated texture to finalize the synthetic sketch. A modified Bayesian classifier is used for classification. The result shows 20% improvement at rank-1 and slightly higher percentage at rank-10. Besides that,