RECOGNIZING FACE SKETCHES BY HUMAN VOLUNTEERS

by

Priyanka Reddy Gangam

Submitted in Partial Fulfillment of the Requirements

for the Degree of

Master of Computing and Information Systems

YOUNGSTOWN STATE UNIVERSITY

August, 2010

RECOGNIZING FACE SKETCHES BY HUMAN VOLUNTEERS

Priyanka Reddy Gangam

I hereby release this thesis to the public. I understand that this thesis will be made available from the OhioLINK ETD Center and the Maag Library Circulation Desk for public access. I also authorize the University or other individuals to make copies of this thesis as needed for scholarly research.

Signature:

Priyanka Reddy Gangam, Student	Date

Approvals:

Yong Zhang, Ph.D., Co-Advisor	Date

John R. Sullins, Ph.D., Co-Advisor	Date

Graciela Perera, Ph.D., Committee Member	Date

Peter J. Kasvinsky, Dean of School of Graduate Studies and Research	Date

DEDICATION

This thesis is dedicated to my father, who taught me that the best kind of knowledge to have is what is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

ABSTRACT

Face sketch recognition by humans has a significant value to both criminal investigators and researchers in computer vision, face biometrics, and cognitive psychology. An important question for both law enforcement agents and scientific researchers is how accurately humans identify hand-drawn face sketches correctly. However, the experimental studies of human performance in recognizing hand-drawn face sketches are still very limited in terms of the number of artists, the number of sketches, and the number of human evaluators involved. In this study, analysis has been concluded based on psychological tests in which 406 volunteers were asked to recognize 250 sketches drawn by 5 different artists. The primary findings are:

i.   The sketch quality has a significant effect on human performance. Inter-artist variation as measured by the mean recognition rate can be as high as 31%.

ii. Participants showed a higher tendency to match multiple sketches to one photo than to second-guess their answers. The multi-match ratio seems correlated to recognition rate, while second-guessing had no significant effect on human performance.

iii. For certain highly recognized faces, their rankings were very consistent using three measuring parameters: recognition rate, multi-match ratio, and second-guess ratio, suggesting that the three parameters could provide valuable information to quantify facial distinctiveness.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Face recognition plays an important role in our daily lives. Human beings have the capability of recognizing faces of parents, friends, family members, teachers, and even strangers. A person gradually develops the ability to identify faces from the time he/she was born. Facial composite images are often used in the criminal investigation process to facilitate the search of someone who has committed a crime. Since the late 19th century, face sketching has been used in criminal investigation and it still continues to be an important forensic technique for law enforcement agencies [1, 2]. In many cases when no traces or evidences are available, drawing a face sketch is the only way to resolve the crime and eyewitness evidence is given considerable weight. Despite its success and popularity, face sketching is known to have uncertainties, which in worse scenario could lead to the false conviction of innocent people [3].

The study of sketch recognition by humans is challenging because a person's cognitive response depends on many factors like distinctiveness of a face, exposure time to a face, the number of sketches used in a test, emotions, motivation, the environment, interest of the participants and the quality of the sketches. All these factors have various degrees of influence on the recognition rate. Therefore, a thorough understanding of how humans identify sketches requires psychological experiments involving a large number of people and a large set of data.

Humans often tend to change their decisions frequently due to the impact of different factors. In sketch recognition test, volunteers showed their hesitation in the form of multi match or second guess [4, 5]. This study addresses three important issues:

- Baseline analysis of the overall performance of all participants.

- Analysis of participants who did multi match (matching the same sketch to various photos or same photo to multiple sketches), and whether their cognitive behaviors reveal more information about facial distinctiveness.

- Analysis of volunteers who did second guess.

In this thesis, the results are reported based on psychological experiments in which 406 volunteers were asked to recognize 250 sketches drawn by 5 different artists. This study has strong implications to automatic face recognition as mentioned in [6], "An understanding of human visual processed involved in face recognition can facilitate and, in turn be facilitated by, better computational models."

2. RELATED WORK

In connection with law and forensic science, face recognition is primarily studied in two communities: computer vision and cognitive psychology. Comprehensive reviews on face perception and recognition can be found in [7, 8, 9, 10, 11, 12]. In computer vision, the research is mainly focused on developing algorithms that search a sketch from a photograph database. Searching was often facilitated by pre-processing methods that bring sketches closer to photographs in a projected space [13]. A transformation function based on the eigen-face method was developed by Tang and Wang [14]. This method enhanced the similarity between sketches and photos. Markov Random Fields model was used to synthesize a sketch from a photo or a photo from a sketch, so that sketch to photo matching can be done in a straight forward manner [15]. The study of using mug-shot database has shown that both legal and global features can improve the searching

accuracy [16]. In cognitive and forensic psychology, Frowd *et al* integrated holistic features into a composite system based on psychological parameters [17]. It was found that facial distinctiveness played an important role in composite matching. Quantifying facial distinctiveness is a promising direction, though designing an objective rating methodology could be challenging. Ideally, a robust distinctiveness rating system should incorporate the performance measures of both algorithms and humans, because human vision may be better equipped to recognize "difficult faces", while algorithms are more efficient in handling a large number of "average" faces [18, 19, 20]. Human vision can be better equipped to recognize "difficult faces", while algorithms are efficient in handling "average faces".

3.   SKETCH DATA

This experiment consists of a data set of 250 hand-drawn sketches. The frontal-view face images, all with neutral face expressions of 50 subjects were selected from a database. These images were printed on white papers (photographs). Five artists were involved in this project. All of them are professor and students of Art. The artists were trained in a workshop on how to draw forensically relevant sketches. For all of the 50 photographs the external features like hair, neck line and ears were removed. All the photographs and sketches were normalized using the eye coordinates. Each artist has drawn 50 sketches, one for each subject. An artist took about 30 to 60 minutes to complete a sketch. All sketches were completed in five month duration. The sketches were divided into two groups randomly: Group A and Group B. Both groups have equal number of sketches. Figure 2.1 shows a photograph and the sketches of that particular

photograph drawn by 5 artists.

Photo	Sketches				
	Artist-1	Artist-2	Artist-3	Artist-4	Artist-5

Figure 1: A face photo and corresponding sketches drawn by five artists.

4. EXPERIMENT DESIGN

4.1 VOLUNTEERS

The experiment was conducted at Youngstown State University. It included a total of 406 volunteers who participated in three tests. Most of the volunteers were students who were taking psychology course at the time of study and the majority of them were freshman or sophomores. The students represented a range of educational backgrounds as they were from different colleges, including College of Liberal Arts and Social Sciences, College of Science, Technology, Engineering, and Mathematics, College of Education, and School of Business. Therefore, the volunteers' performance can be considered representative of a much larger population.

4.2 EXPERIMENT DESIGN AND PROTOCOL

The experiments were conducted in three different sessions. The first session had 61 volunteers, the second session had 184 volunteers, and the third session had 161 volunteers making a total of 406 volunteers during the entire testing period.

At the beginning of each of the three sessions, the researcher explained the nature and scope of the experiments as well as the risk and benefits of participating in the study.

4

All of the volunteers signed an IRB (Institutional Review Board) approved consent form as they were given credit for participating in the experiment. During a session, each volunteer was given two papers: a reference sheet and an answer sheet. The reference sheet had the photographs of the subjects. In the first session a reference sheet had 50 photographs and 25 photographs in the second and third sessions. Each photograph corresponded to a different person. There were no duplicates of the subjects. Each photograph was numbered. The answer sheet had sketches of the subjects. An answer sheet contained 50 sketches in the first session and 25 sketches in the second and third sessions. Each sketch a answer sheet corresponded to a different person. All sketches in an answer sheet were drawn by the same artist. Both reference and answer sheets contained the photographs and sketches of the same 25 or 50 persons, but their positions were randomized. Examples of reference sheet and answer sheet are shown in figure 2.

Figure 2: Examples of reference sheet (photos, left) and answer sheet (sketches, right).

For each sketch, participants were asked to find a match in the reference sheet. Two special cases were examined in the experiment: Multi-match and second guessing. In multi-match, a sketch can be matched to multiple photos or the photos can be matched to multiple sketches. The multi-match was only considered during the second and third sessions. For second guessing, the volunteers were informed that, for making a second guess, they need to mark out the original photo ID on the answer sheet and then write down the new photo ID. So, the result of second match can be counted later. As the first session had twice the number of photos/sketches (50) than that in the second and third sessions (25), it took 30 to 60 minutes to complete the task, while the second and third sessions took less than 20 minutes. Light refreshments were also provided to all the volunteers during the first session to minimize the impact of fatigue. The complete information of volunteers and the test conditions are summarized in Table 1.

Table 1: Information of Volunteers and Test Conditions

	First Session	**Second Session**	**Third Session**
Num. of Volunteers	61	184	161
Num. of Sketches	50	25	25
Multiple Matches	N/A	Allowed	Allowed
Second Guess	Marked	Marked	Marked
Refreshments	Yes	No	No
Finishing Time (min)	30 - 60	10 -20	10 – 20

4.3 DATA PROCESSING

The answer sheets were collected at the end of each session. The answer sheets were sorted into their respective groups and numbered. A spreadsheet has been used in which the photo IDs or the subject of the sketches (*s01, s02, ..., sN*) were listed in the first column and the volunteer numbers (*st1, st2, ..., stM*) were listed in the first column. This way, results can be tabulated into a matrix and the answers can be counted easily. Recognition accuracies of each photo and each student were listed in the last column and last row respectively. Use the data, five tasks were performed:

- Task-1: During this task, if a student matched the sketch correctly to its respective photograph then a '1' is entered in the corresponding cell. Otherwise a '0' is entered. A value of '1' indicates a correct match and '0' indicates a wrong answer. The numbers of correct and wrong matches were counted for each volunteer and each photograph. Figure 3 illustrates the spreadsheet for Task 1.

- Task-2: Some of the volunteers did not provide their answer for a few sketches. In this case, the cell of unanswered sketches was marked by red color (Figure 4).

- Task-3: As mentioned earlier, the volunteers were allowed to second guess. This task counts the number of second-guessed sketches. In case a volunteer scribbled over old answers, we only used the best one. The cells of green color in Figure 5 correspond to the ones that were second-guessed.

- Task-4: This task accounts the one-to-many matches. Only a few volunteers matched the same photograph to multiple sketches. The answers marked in yellow color along with the IDs are shown in Figure 6. If it was a correct match, the font color was changed to red.

- Task-5: This task handles many-to-one case: matching the same sketch to multiple photographs. This task was very time consuming because many volunteers had this type of match. Those answers were marked with blue color (Figure 7).

Sketch 1 Group A

	st1	st2	st3	st4	st5	st6	st7	st8	st9	st10	st11	st12	st13	st14	st15	st16	st17	st18	st19	st20	# of correct match	correct ratio	# of students	statistics by subject	
s01	0	0	0	0	0	1	0	0	0	0	0	0	0	0							1	0.06	18	min	0.00
s02	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00		max	0.61
s03	0	0	1	0	0	0	0	0	0	0	0	0	1	0							2	0.11		mean	0.24
s04	1	0	0	1	0	0	1	0	1	1	1	1	0	1							8	0.44		std	0.20
s05	1	0	0	0	1	0	1	1	1	1	0	0	0	1							7	0.39			
s06	0	0	0	1	1	0	1	0	0	1	1	0	1	1							7	0.39			
s07	1	0	1	0	1	1	0	0	1	1	0	1	0	0							7	0.39			
s08	1	0	0	1	1	1	0	0	1	1	0	0	0	1							7	0.39			
s09	0	0	0	1	1	1	1	0	0	0	1	0	0	0							5	0.28			
s10	1	0	1	1	1	1	1	1	1	1	1	0	0	1							11	0.61			
s11	1	1	0	1	1	0	1	1	1	1	1	0	1	1							11	0.61			
s12	1	0	0	0	0	0	1	0	1	1	0	0	1	0							6	0.33			
s13	1	0	0	0	0	1	1	0	0	1	1	1	0	1							7	0.39			
s14	0	0	0	0	0	0	0	0	1	0	0	0	1	0							2	0.11			
s15	1	0	0	0	0	0	1	0	0	1	0	0	0	0							3	0.17			
s16	0	0	0	1	0	0	1	0	0	0	0	0	0	0							2	0.11			
s17	0	0	0	1	0	1	0	0	0	0	0	0	0	0							2	0.11			
s18	1	1	1	1	1	1	0	0	1	1	0	1	1	1							11	0.61			
s19	0	1	1	1	1	1	0	1	0	0	0	0	0	0							6	0.33			
s20	0	0	0	0	0	0	0	0	0	1	0	0	0	0							1	0.06			
s21	0	0	0	0	1	0	0	0	0	0	0	0	0	0							1	0.06			
s22	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00			
s23	0	0	0	0	0	0	0	0	1	0	0	0	0	0							1	0.06			
s24	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00			
s25	0	0	0	0	1	0	0	0	0	1	0	0	0	0							2	0.11			
# of correct match	10	3	5	10	10	10	10	4	9	13	7	4	6	9	0	0	0	0	0	0					
recognition rate	0.40	0.12	0.20	0.40	0.40	0.40	0.40	0.16	0.36	0.52	0.28	0.16	0.24	0.36	0.00	0.00	0.00	0.00	0.00	0.00					

statistics per student:

min	max	mean	std
0.00	0.52	0.24	0.17

Figure 3: Spread sheet for Task 1.

Sketch 1 Group A

	st1	st2	st3	st4	st5	st6	st7	st8	st9	st10	st11	st12	st13	st14	st15	st16	st17	st18	st19	st20	# of correct match	correct ratio	# of students	statistics by subject	
s01	0	0	0	0	0	1	0	0	0	0	0	0	0	0							1	0.06	18	min	0.00
s02	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00		max	0.61
s03	0	0	1	0	0	0	0	0	0	0	0	0	1	0							2	0.11		mean	0.25
s04	1	0	0	1	0	0	1	0	1	1	1	1	0	1							8	0.44		std	0.20
s05	1	0	0	0	1	0	1	1	1	1	0	0	0	1							7	0.39			
s06	0	0	0	1	1	0	1	0	0	1	1	0	1	1							7	0.39			
s07	1	1	1	0	1	1	0	0	1	1	0	1	0	0							8	0.44			
s08	1	0	0	1	1	1	0	0	1	1	0	0	0	1							7	0.39			
s09	0	0	0	1	1	1	1	0	0	0	1	0	0	0							5	0.28			
s10	1	0	1	1	1	1	1	1	1	1	1	0	0	1							11	0.61			
s11	1	1	0	1	1	0	1	1	1	1	1	0	1	1							11	0.61			
s12	1	0	0	0	0	0	1	0	1	1	0	0	1	1							6	0.33			
s13	1	0	0	0	0	1	1	0	0	1	1	1	0	1							7	0.39			
s14	0	0	0	0	0	0	0	0	1	0	0	0	1	0							2	0.11			
s15	1	0	0	0	0	0	1	0	0	1	0	0	0	0							3	0.17			
s16	0	0	0	1	0	0	1	0	0	0	0	0	0	0							2	0.11			
s17	0	0	0	1	0	1	0	0	0	0	0	0	0	0							2	0.11			
s18	1	1	1	1	1	1	0	0	1	1	0	1	1	1							11	0.61			
s19	0	1	1	1	1	1	0	1	0	0	0	0	0	0							6	0.33			
s20	0	0	0	0	0	0	0	0	0	0	1	0	0	0							1	0.06			
s21	0	0	0	0	0	1	0	0	0	0	0	0	0	0							1	0.06			
s22	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00			
s23	0	0	0	0	0	0	0	0	0	1	0	0	0	0							1	0.06			
s24	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00			
s25	0	0	0	0	1	0	0	0	0	1	0	0	0	0							2	0.11			
# of correct match	10	4	5	10	10	10	10	4	9	13	7	4	6	9	0	0	0	0	0	0					
recognition rate	0.40	0.16	0.20	0.40	0.40	0.40	0.40	0.16	0.36	0.52	0.28	0.16	0.24	0.36	0.00	0.00	0.00	0.00	0.00	0.00					

statistics per student:

min	max	mean	std
0.00	0.52	0.25	0.17

Figure 4: Spread sheet for Task 2.

Figure 5: Spread sheet for Task 3.

Sketch 1   Group A

|	st1	st2	st3	st4	st5	st6	st7	st8	st9	st10	st11	st12	st13	st14	st15	st16	st17	st18	st19	st20	# of correct match	correct ratio	# of students	statistics by subject	
s01	0	0	0	0	0	1	0	0	0	0	0	0	0	0							1	0.06	18	min	0.00
s02	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00		max	0.61
s03	0	0	1	0	0	0	0	0	0	0	0	0	1	0							2	0.11		mean	0.25
s04	1	0	0	1	0	0	1	0	1	1	1	1	0	1							8	0.44		std	0.20
s05	1	0	0	0	1	0	1	1	1	1	0	0	0	1							7	0.39			
s06	0	0	0	1	1	0	1	0	0	1	1	0	1	1							7	0.39			
s07	1	1	1	0	1	1	0	0	1	1	0	1	0	0							8	0.44			
s08	1	0	0	1	1	1	0	0	1	1	0	0	1	0							7	0.39			
s09	0	0	0	1	1	1	1	0	0	0	1	0	0	0							5	0.28			
s10	1	0	1	1	1	1	1	1	1	1	1	0	1	1							11	0.61			
s11	1	1	0	1	1	0	1	1	1	1	1	0	1	1							11	0.61			
s12	1	0	0	0	0	0	1	0	1	1	0	0	1	1							6	0.33			
s13	1	0	0	0	0	1	1	0	0	1	1	1	0	1							7	0.39			
s14	0	0	0	0	0	0	0	1	0	0	0	1	0								2	0.11			
s15	1	0	0	0	0	0	1	0	0	1	0	0	0	0							3	0.17			
s16	0	0	0	1	0	0	1	0	0	0	0	0	0	0							2	0.11			
s17	0	0	0	1	0	1	0	0	0	0	0	0	0	0							2	0.11			
s18	1	1	1	1	1	1	0	0	1	1	0	1	1	1							11	0.61			
s19	0	1	1	1	1	1	0	1	0	0	0	0	0	0							6	0.33			
s20	0	0	0	0	0	0	0	0	0	1	0	0	0	0							1	0.06			
s21	0	0	0	0	0	1	0	0	0	0	0	0	0	0							1	0.06			
s22	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00			
s23	0	0	0	0	0	0	0	0	0	1	0	0	0	0							1	0.06			
s24	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00			
s25	0	0	0	0	1	0	0	0	0	1	0	0	0	0							2	0.11			
# of correct match	10	4	5	10	10	10	10	4	9	13	7	4	6	9	0	0	0	0	0	0					
recognition rate	0.40	0.16	0.20	0.40	0.40	0.40	0.40	0.16	0.36	0.52	0.28	0.16	0.24	0.36	0.00	0.00	0.00	0.00	0.00	0.00					

statistics per student: min max mean std → 0.00 0.52 0.25 0.17

Figure 6: Spread sheet for Task 4.

	st1	st2	st3	st4	st5	st6	st7	st8	st9	st10	st11	st12	st13	st14	st15	st16	st17	st18	st19	st20	# of correct match	correct ratio	# of students	statistics by subject	
s01	0	0	0	0	0	1	0	0	0	0	0	0	0	0							1	0.06	18	min	0.00
s02	0	0	0	0	0	0	0	0	0	0	0	0	0	0							0	0.00		max	0.61
s03	0	0	1	0	0	0	0	0	0	0	0	0	1	0							2	0.11		mean	0.24
s04	1	0	0	1	0	0	1	0	1	1	1	1	0	1							8	0.44		std	0.20
s05	1	0	0	0	1	0	1	1	1	1	0	0	0	1							7	0.39			
s06	0	0	0	1	1	0	1	0	0	1	1	0	1	1							7	0.39			
s07	1	1	1	0	1	1	0	0	1	1	0	1	0	0							8	0.44			
s08	1	0	0	1	1	1	0	0	1	1	0	0	1								7	0.39			
s09	0	0	0	1	1	1	1	0	0	0	1	0	0	0							5	0.28			
s10	1	0	1	1	1	1	1	1	1	1	1	0	0	1							11	0.61			
s11	1	1	0	1	1	0	1	1	1	1	1	0	1	1							11	0.61			
s12	1	0	0	0	0	0	1	0	1	1	0	0	12/14								5	0.28			
s13	1	0	0	0	0	1	1	0	0	1	1	1	0	0							7	0.39			
s14	0	0	0	0	0	0	0	1	0	0	0	1	0								2	0.11			
s15	1	0	0	0	0	0	1	0	0	1	0	0	0	0							3	0.17			
s16	0	0	0	1	0	0	1	0	0	0	0	0	0								2	0.11			
s17	0	0	0	1	0	0	0	0	0	0	0	0	0	0							2	0.11			
s18	1	1	1	1	1	1	0	0	1	1	0	1	1	1							11	0.61			
s19	0	1	1	1	1	1	0	1	0	0	0	0	0	0							6	0.33			
s20	0	0	0	0	0	0	0	0	0	1	0	0	0	0							1	0.06			
s21	0	0	0	0	0	1	0	0	0	0	0	0	0								1	0.06			
s22	0	0	0	0	0	0	0	0	0	0	0	0	0								0	0.00			
s23	0	0	0	0	0	0	0	0	0	1	0	0	0								1	0.06			
s24	0	0	0	0	0	0	0	0	0	0	0	0	0								0	0.00			
s25	0	0	0	0	1	0	0	0	0	1	0	0	0	0							2	0.11			
# of correct match	10	4	5	10	10	10	10	4	9	13	7	4	6	8	0	0	0	0	0	0					
recognition rate	0.40	0.16	0.20	0.40	0.40	0.40	0.40	0.16	0.36	0.52	0.28	0.16	0.24	0.32	0.00	0.00	0.00	0.00	0.00	0.00					

statistics per student: min max mean std → 0.00 0.52 0.24 0.17

Figure 7: Spread sheet for Task 5.

	st1	st2	st3	st4	st5	st6	st7	st8	st9	st10	st11	st12	st13	st14	st15	st16	st17	st18	st19	st20	# of correct match	correct ratio	# of students	statistics by subject	
s01	0	0	10	0	0	1	0	0	0	0	0	8	0	22							41	2.28	18	min	0.06
s02	0	4	17	0	0	0	0	8	0	0	0	0	0	0							29	1.61		max	4.00
s03	0	0	1	15	0	0	0	0	0	0	10	1	0								27	1.50		mean	2.07
s04	1	0	0	1	21	0	1	0	1	1	1	4	0	4							35	1.94		std	1.03
s05	1	0	0	25	1	0	1	1	1	5	0	0	11	1							47	2.61			
s06	0	0	0	1	6	20	1	0	0	1	1	0	1	6							37	2.06			
s07	1	4	1	18	1	1	0	0	1	1	0	1	0	0							29	1.61			
s08	1	0	0	8	8	1	0	0	1	1	0	10	0	1							31	1.72			
s09	0	0	15	9	1	1	1	0	0	5	1	0	0	0							33	1.83			
s10	1	0	10	1	1	1	10	1	1	1	1	0	0	10							38	2.11			
s11	1	1	0	1	1	0	1	1	1	1	1	20	11	1							41	2.28			
s12	1	0	0	0	6	0	1	0	1	1	0	0	1	1							12	0.67			
s13	1	0	17	8	8	1	1	8	0	13	1	1	0	13							72	4.00			
s14	0	0	15	15	0	0	0	0	1	0	10	1	6								48	2.67			
s15	1	0	0	25	0	0	1	0	0	1	0	0	17								45	2.50			
s16	0	0	0	1	0	19	1	0	0	0	0	4	0	22							47	2.61			
s17	0	19	0	17	0	1	0	0	0	0	0	0	0								37	2.06			
s18	1	1	1	18	1	1	0	1	1	0	18	1	1								45	2.50			
s19	0	19	1	1	1	19	0	1	0	0	20	0	4								66	3.67			
s20	0	0	0	3	0	0	0	0	0	0	1	0	0	0							4	0.22			
s21	0	0	0	3	0	1	0	0	0	0	0	0	0								4	0.22			
s22	0	0	0	17	0	0	0	0	0	13	0	8	0	10							48	2.67			
s23	0	0	0	0	0	0	0	0	0	1	0	0	0								1	0.06			
s24	0	0	0	0	24	0	0	0	0	0	0	0	17								41	2.28			
s25	0	0	0	9	1	20	10	0	0	1	0	18	13								72	4.00			
# of correct match	10	48	88	196	81	87	29	20	9	47	7	132	27	149	0	0	0	0	0	0					
recognition rate	0.40	1.92	3.52	7.84	3.24	3.48	1.16	0.80	0.36	1.88	0.28	5.28	1.08	5.96	0.00	0.00	0.00	0.00	0.00	0.00					

statistics per student: min max mean std → 0.00 7.84 2.07 2.35

9

4.4 EVALUATION METHODS

The results of all three sessions were tabulated into a 2D score matrix as shown in Table 2. The total value of the numbers (either '0' or '1') along the rows and columns was calculated and used to compute two recognition rates:

$$R_S = H_{row}/N,$$

$$R_v = H_{col}/M,$$

where H_{row} is the number of correct answers in each row (the number of correct matches by all volunteers with respect to a sketch), H_{col} is the number of correct answers in each column (the number of correct matches by a volunteer for all sketches), N is the number of volunteers and M is the number of sketches involved in a test.

R_S represents the recognition rate of a sketch averaged over all volunteers, and R_V represents the recognition rate of the volunteer averaged over all sketches. In this case, R_S was used to characterize the subjects in sketches while R_V was used to analyze the performance variation of the volunteers. For example, which face portrayed in sketches was the most or least recognizable? The ranking of faces in terms of their distinctiveness can then be quantified. During the second guess computing process, only the cells corresponding to a second guess or the ones in green color mark in the spreadsheet were counted. Similarly, for multi-match computation, only the cells corresponding to multi-match were counted.

Human performance was assessed using statistical inference methods that have been accepted widely accepted in experimental psychology, including ANOVA and post hoc hypothesis tests.

Table 2: The Matrix for Calculating Recognition Rates (R_V, R_S)

	Volunt. 1	Volunt. 2	…	Volunt. N	# of Hits (H_{row})	Recog. Rate (R_s)
Sketch-1	1	0	…	0	26	26/N
Sketch-2	0	0	…	1	31	31/N
.						
Sketch-M	0	1	…	0	28	28/N
# of Hits (H_{col})	17	24	…	21		
Recog. Rate (R_v)	17/M	24/M	…	21/M		

5. RESULTS AND DISCUSSION

5.1 BASELINE PERFORMANCE

Three issues have been addressed in the baseline analysis: (i) Does the sketch recognition rate change significantly from one artist to another? If so, which artist had a better or worse performance? (ii) Does the number of sketches used in a test have a systematic impact on recognition rate? (iii) Which group of faces in sketches is more recognizable? Table 3 shows the statistics of the recognition rate (R_V) of first session and Table 4 shows the statistics of recognition rate (R_V) of the second and third sessions. Figure 8 shows the mean recognition rates of five artists. It is clear that Artists-4 has the highest recognition rates followed by Artist-2 and Artist-5. The rate of using 50 sketches was lower than that of using 25 sketches, though the relative performances of five artists remained same.

11

Table 3: Recognition Statistics of the 1st Session

Recognition Rate (R_v)	Artist 1	Artist 2	Artist 3	Artist 4	Artist 5
Mean	30%	48%	32%	61%	47%
Min	6%	16%	13%	30%	18%
Max	42%	72%	49%	82%	76%

Table 4: Recognition Statistics of the 2nd and 3rd Sessions

Recognition Rate (R_v)	Artist 1	Artist 2	Artist 3	Artist 4	Artist 5
Mean	43%	62%	48%	66%	57%
Min	16%	32%	12%	28%	20%
Max	88%	100%	92%	92%	80%
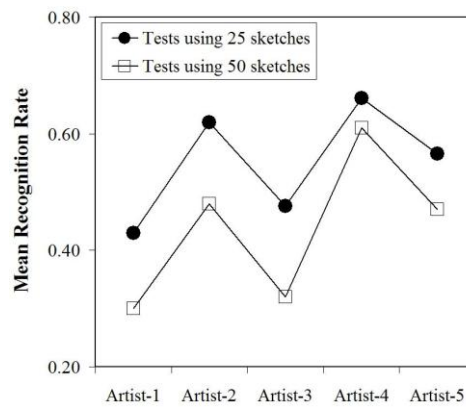
Figure 8: The mean recognition rates (R_v) of the three sessions

Two one-way ANOVA were conducted, considering artist as the categorical factor and R_V as the dependent variable. First ANOVA was conducted on the data of the first session which included 50 sketches and 61 volunteers. The Levene test suggested no violation of homogeneity of variance assumption across groups: $F(4, 56) = 1.93$, $p = 0.12$. The differences among the five artists were significant: $F(4,56) = 7.57$, $p < 0.001$, which had an effect size of $\eta^2 = 0.35$. The Tukey's HSD (Honestly Significant Differences) test ($\alpha = 0.05$) indicated that Artist-1 scored significantly lower than Artist-2, Artist-4 and Artist-5, while Artist-4 was also better than Aritist-3.

The second ANOVA was conducted on the data of the second and third sessions which included 345 volunteers and 25 sketches. Here also no violation of homogeneity variance was found by the Levene test: $F(4, 340) = 1.14$, $p = 0.34$. Again there was a significant variation among artists: $F(4, 340) = 28.99$, $p < 0.01$, with an effect size of $\eta^2 = 0.25$. The pair wise comparison of Tukey's test ($\alpha = 0.05$) generated similar relative rankings of artists as in the first ANOVA.

Therefore, from the above data two things can be concluded: (i) the sketches of Artist-1 are most challenging with the average recognition rate of 30% for the first session and the average recognition rate of 43% for the second session; (ii) the sketches of Artist-4 are of high quality with the recognition rate of 61% for the first session and 66% for the second session. Artist 4 was an Art Professor and all the other artists were students which imply that experience and training play an important role in constructing reliable sketch faces.

Selecting the appropriate number of sketches for a test is extremely important to reveal the true cognitive capability of participants, as it is related to the fatigue effect that

could undermine the internal validity of a psychological experiment [23]. As the number of sketches increases the performance of the participants could decrease due to the loss of concentration and interest. But if reduce the number of sketches or use a very small number of sketches, the recognition rate could be so high that suppresses the cognitive diversity of the participants. Figure 9 shows the correlation of recognition rates R_S between the test using 25 sketches and the test using 50 sketches. A good correlation between the two designs ($r = 0.84$, $p < 0.01$) and their balanced recognition rates suggest that using $20 - 50$ sketches is a reasonable choice.
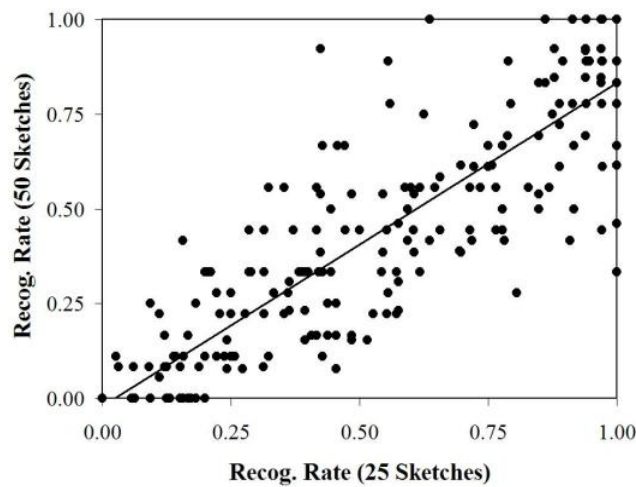
Figure 9: The correlation of recognition rates (R_s) between the test 25- 50 sketches.

In a recent study of computer generated composites, it was found that facial distinctiveness is an important factor in target naming [18]. From the neurobiological findings it was found that certain neurons are more responsive to a particular facial feature [24]. It is important to find facial features that are more distinctive. Therefore a rating system that can quantify the distinctiveness of a face sketch which takes into

14

account the performance measures of a large number of human evaluators and multiple algorithms is devised. Figure 10 shows the six most and least recognized faces as ranked by their recognition rates (R_S) averaged over five artists.

The highly recognized faces had a few characteristics: (i) Older age groups (two subjects of >40 years old and one subject of >60 years old); (ii) Having hair and mustaches; (iii) Minority ethnic groups (one Asian female and one African American male); (iv) Having facial expressions. Human vision is more responsive to salient visual stimuli. It has been found that smiling faces are better for recognition [25]. Young faces and faces of clean and of neutral expression are the least recognized ones.
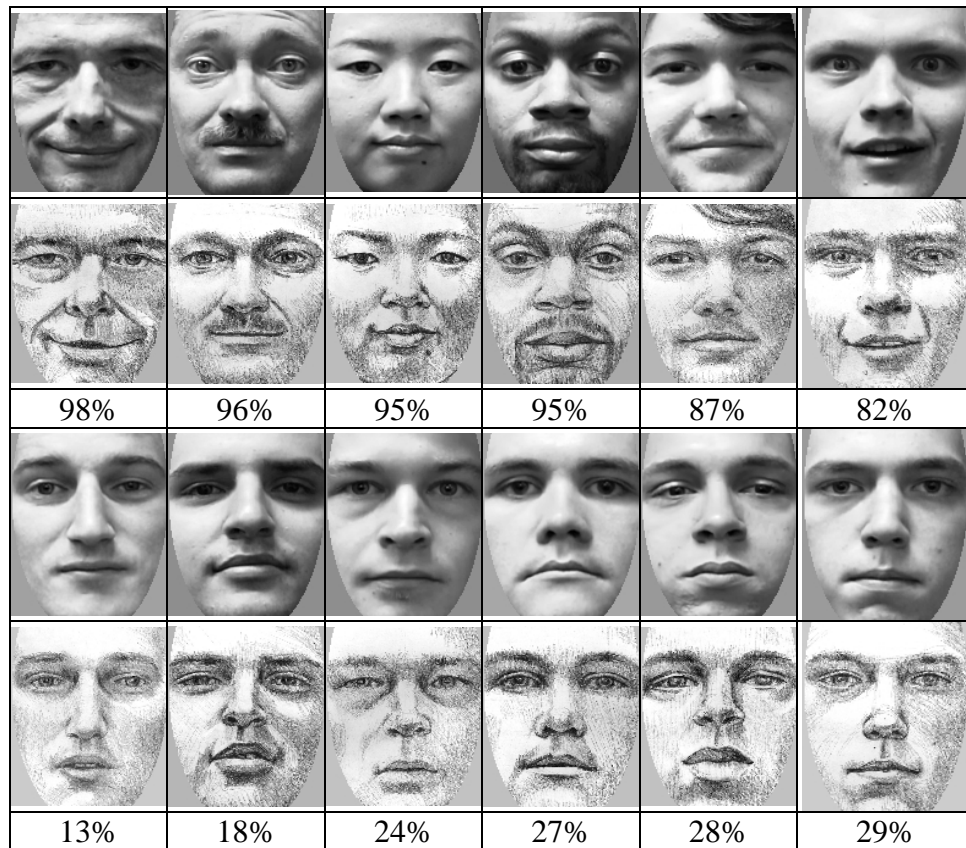
Figure 10: The most and least recognized faces as ranked by their recognition rates (R_s*).*

5.2 MULTI-MATCH

Usually in forensic investigations the eyewitnesses are asked to identify the suspect from a group of people, in the form of either mug shots or face composites. Eyewitness interview could lead to a false identification because of the large uncertainties caused by several factors, including the distinctiveness and typicality of a suspect's face [26, 3]. In sketch recognition, a similar problem occurred when the volunteers found that a particular sketch looks like multiple photos or *vice versa*. Multi-match could be a useful measure of facial distinctiveness that can be tackled in developing better recognition algorithms or face composite systems. Multi match can be of two types: (i) One sketch to multiple photos; (ii) One photo to multiple sketches.

In this research, less than 2% of volunteers did the first type of multi-match, so focus was mainly on the second type. Table 5 shows the statistics of multi-match for second and third sessions. Multi-match ratio is the percentage of volunteers who did multi-match to a particular sketch. Almost all the artists had the same multi-match ratio except Artist-4 with just 8%. This clearly shows that the volunteers could recognize the sketches easily that has good quality. Figure 11 shows the mean recognition rates (R_V). The volunteers who did multi-match had lower recognition rates than those who did not do multi-match.

A two-way ANOVA test was also conducted. The R_V values of 345 volunteers from the second and third sessions were subjected to a 5x2 factorial ANOVA with artist and multi-match as between subject factors. Levene test indicated no violation of the homogeneity of variance assumption: $F(9, 335) = 0.84$, p=0.58. The main effect of artist was significant: $F(4, 335) = 27.72$, $p < 0.001$, $\eta^2 = 0.25$, similar to the one-way ANOVA

16

in the baseline case. The multi-match effect was significant: $F(1, 335) = 20.55$, $p < 0.001$, but with a much smaller effect size of $\eta^2 = 0.06$. The interaction failed to attain significance: $F(4, 335) = 0.75$, $p = 0.56$. The Tukey's test on artist generated similar outcomes as in the baseline case. So, the ANOVA was consistent with the observations made from Table 5 and Figure 11.

Table 5: Multi-Match Statistics of 2nd and 3rd Sessions

	Artist 1	Artist 2	Artist 3	Artist 4	Artist 5
Mean recog. rate of volunteers who **did not** do multi-match.	48%	69%	51%	69%	59%
Mean recog. rate of volunteers who **did** multi-match.	40%	57%	45%	62%	55%
Multi-match ratio.	16%	14%	16%	8%	15%
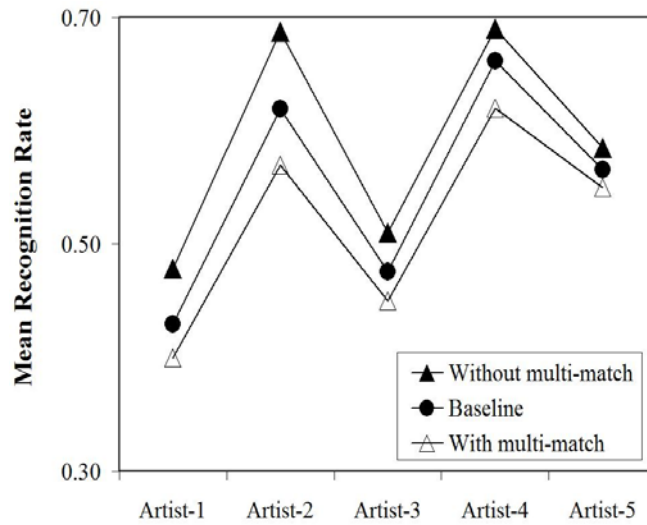
Figure 11: Comparison of volunteers who did multi-match versus those who did not, and the baseline performance.

Figure 12 shows the multi-match ratios averaged over five artists plotted against the mean recognition rates of R_S. The multi-match ratio was correlated with the recognition rate ($r = 0.79$ $p = 0.01$). The faces were ranked based on their multi-match ratios and Figure 13 shows a few representative ones that were picked from the top ten and bottom ten, respectively. From Figure 13 and Figure 10, it is clear that the six faces that received the fewest multi-match hits were also the highly recognized in the baseline case, with a slightly different order. The faces of higher multi-match ratios also correspond well to the least recognized faces which suggest that multi match ratio could be an important measuring parameter of facial distinctiveness.
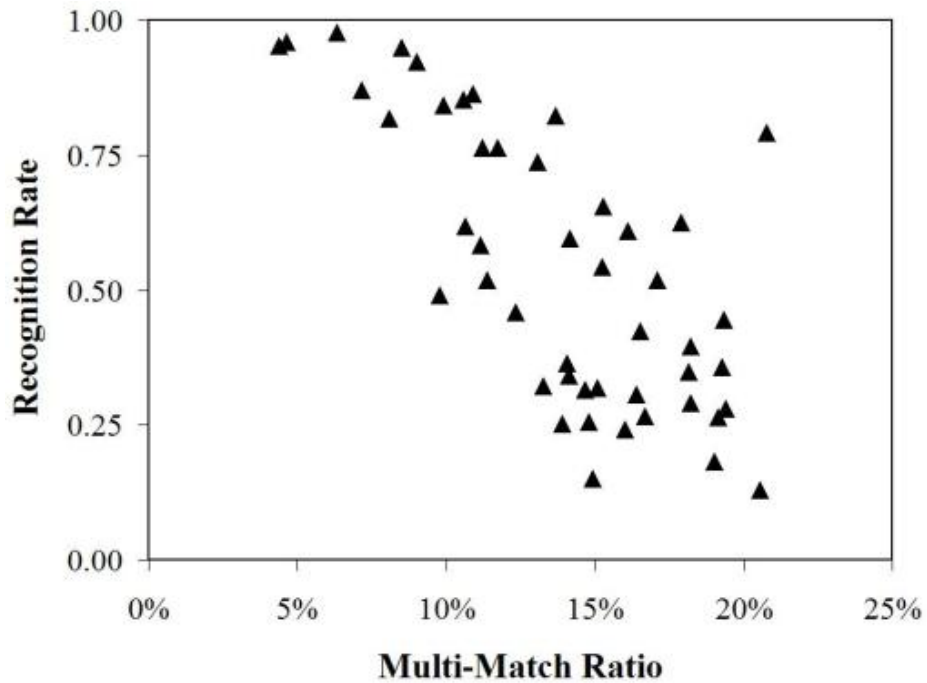
Figure 12: Relationship between recognition rate (R_S*) and multi-match ratio.*
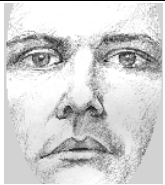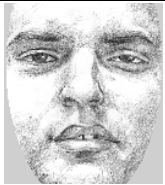
4%	5%	6%	7%	8%	8%

21%	19%	19%	19%	18%	18%

Figure 13: The least and most multi-matched faces as ranked by their multi-match ratios.

5.3 SECOND GUESSING

Second guessing was usually beneficial to people who changed their test answers, because many complex thinking processes were involved. Here in the sketch recognition tests, it is important to know whether second guessing has any positive or negative impact on the participants' recognition rates.

Second guessing statistics of second and third sessions are shown in Table 6. The

second-guess ratio is defined as the percentage of volunteers who second-guessed a particular sketch. A two-way ANOVA with a 5x2 fully factorial design was carried out. Artist and second-guess served as two factors, and R_V of the second and third sessions with 345 volunteers was the dependent variable. No violation of homogeneity of variance assumption was found in Levene test: $F(9, 335) = 0.98$, $p = 0.46$. The main effect of artist was significant: $F(4, 335) = 28.93$, $p<0.001$, $\eta^2 = 0.26$. But the main effect of second-guess fell short of significance: $F(1,335)=2.48$, $p = 0.12$, and no significant interaction was found: $F(4, 335) = 0.72$, $p = 0.58$.

After a thorough statistical analysis it was found that second guessing did not help the volunteers' overall performance because only a small percentage of sketches were second guessed. From Table 5 and Table 6, it is clear that the second-guess ratios were much lower than that of the multi match ratios. Because of the lack of correlation between the second-guess ratio and the multi-match ratio, it implies that the thinking processes involved are probably independent of each other.

No correlation was observed between the second guess ratio and the recognition rate of R_S. Figure 14 shows the subjects ranked based on their second guessed ratios. From Figure 14 and Figure 10, it is clear that least second guessed faces matched well with highly recognized faces in baseline analysis. But the faces that received the most second guess hits are quite different from least recognized faces. This implies that the second-guess ratio may still be considered in facial distinctiveness calculation, but probably only for the less second-guessed faces.

Table 6:Second-Guess Statistics of 2ⁿᵈ and 3ʳᵈ Sessions

	Artist 1	Artist 2	Artist 3	Artist 4	Artist 5
Mean recog. rate of volunteers who **did not** do second-guess.	40%	62%	48%	65%	53%
Mean recog. rate of volunteers who **did** second-guess.	45%	62%	47%	67%	60%
Second-guess ratio.	5%	4%	4%	5%	4%
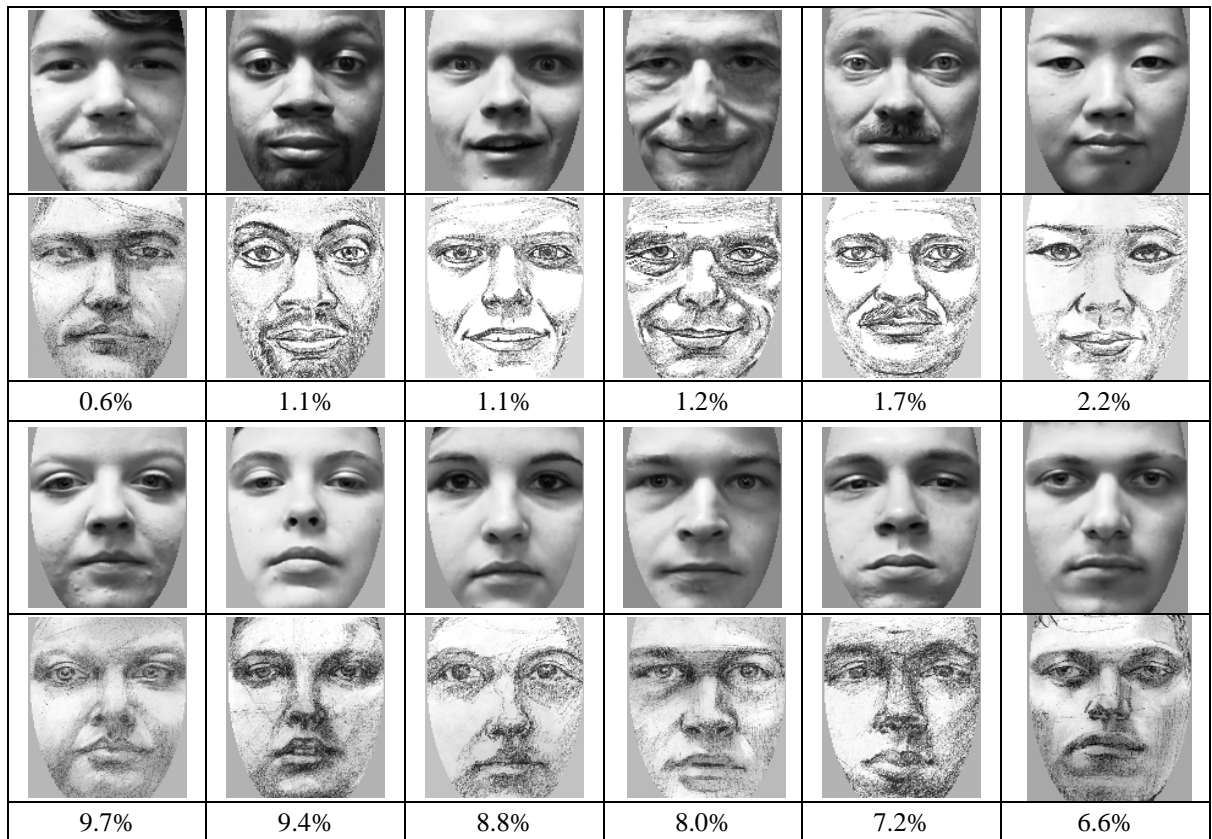
Figure 14: The most and least second-guessed faces as measured by their second-guess ratios.

6. CONCLUSIONS

This thesis is an attempt to understand how humans recognize a person from hand-drawn face sketches. It presents a sketch recognition study involving a large number of volunteers and sketches drawn by various artists. The findings of this study have strong implications to algorithm based automatic face recognition, pattern recognition and biometrics in general. The major findings are summarized below:

- Based on statistical analysis, it was found that the sketch quality had a significant impact on human recognition rate. The baseline analysis shows that higher recognition rates can be achieved by improving the sketch quality.

- Based on the recognition rate, multi-match ration or second-guess ratio, some faces were ranked as highly recognizable. The high consistency suggests that those particular faces possess unique features that make them distinct. Recognition rate, multi-match ratio and second-guess ratio are key parameters to be considered for an objective rating system of facial distinctiveness.

- More number of sketches have been multi-matched than being second-guessed.

- Correlation was observed between the multi-match ratio and the recognition rate.

- No correlation was observed between second-guess ratio and recognition rate.

- It is advisable to have 25 to 50 sketches for future investigations with a similar experimental protocol.

7. REFERENCES

[1] K. T. Taylor, Forensic Art and Illustration, CRC Press, 2000.

[2] D. Mcquiston-Surrett, L. D. Topp, and R. S. Malpass, "Use of facial composite systems in US law enforcement agencies", Psychology, Crime and Law, 12(5), pp. 505-517, 2006.

[3] G. L. Wells and L. E. Hasel, "Facial composite production by eyewitnesses", Current Directions in Psychological Science, 16(1), pp. 6-10, 2007.

[4] J. Kruger, D. Wirtz and D. T. Miller, "Counterfactual thinking and the first instinct fallacy", Journal of Personality and Social Psychology, 88(5), pp. 725-735, 2005.

[5] L. T. Benjamin, T. A. Cavell, and W. R. Shallenberger, "Staying with the initial answers on objective tests: Is it a myth?" Teaching of Psychology, 11(3), pp. 133-141, 1984.

[6] P. Sinha, B. J. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: 19 results all computer vision researchers should know about", Proceedings of the IEEE, 94(11), pp. 1948-1962, 2006.

[7] V. Bruce, Recognizing Faces, London: Lawrence Erlbaum Associates, 1988.

[8] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," ACM Computing Surveys, 35(4), pp. 399-458, 2003.

[9] S. Z. Li and A. K. Jain (editors), Handbook of Face Recognition, Springer, 2005.

[10] K. W. Bowyer, K. Chang, and P. J. Flynn. "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," Computer Vision and Image Understanding, 101(1), pp. 1-15, 2006.

[11] H. Wechsler, Reliable Face Recognition Methods, System Design, Implementation and Evaluation, Springer, 2007.

[12] A. A. Ross, K. Nandakumar, and A. K. Jain, Handbook of Multibiometrics, Springer, 2006.

[13] R. G. Uhl, and N. V. Lobo, "A framework for recognizing a facial image from a police sketch", Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 586-593, San Francisco, 1996.

[14] X. Tang, and X. Wang, "Face sketch recognition", IEEE Transactions on Circuits and Systems for Video Technology, 14(1), pp. 50-57, 2004.

[15] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(11), pp. 1955-1967, 2009.

[16] P. C. Yuen, and C. H. Man, "Human face image searching system using sketches", IEEE Transactions on Systems, Man, and Cybernetics, Part A, 37(4), pp. 493-504, 2007.

[17] C. D. Frowd, V. Bruce, A. McIntyre, D. Ross, S. Fields, Y. Plenderleith, and P. J. B. Hancock, "Implementing holistic dimensions for a facial composite system", Journal of Multimedia, 1(3), pp. 42-51, 2006.

[18] C. D. Frowd, D. Carson, H. Ness, J. Richardson, L. Morrison, S. McLanaghan, and P. J. B. Hancock, "A forensically valid comparison of facial composite systems",

Psychology, Crime & Law, 11(1), pp. 33-52, 2005.

[19] C. D. Frowd, D. McQuiston-Surrett, S. Anandaciva, C. E. Ireland, and P. J. B. Hancock, "An evaluation of US systems for facial composite production", *Ergonomics*, 50, pp. 562-585, 2007.

[20] C. D. Frowd, D. Carson, H. Ness, D. McQuiston, J. Richardson, H. Baldwin, and P. J. B. Hancock, "Contemporary Composite Techniques: the impact of a forensically-relevant target delay", *Legal & Criminological Psychology*, 10, pp. 63-81, 2005.

[21] A. J. O'Toole, P. J. Phillips, F. Jiang, J. H. Ayyad, N. Penard, and H. Abdi, "Face recognition algorithms surpass humans matching faces over changes in illumination", IEEE Trans. on Pattern Analysis and Machine Intelligence,29(9), pp. 1642-1646, 2007.

[22] A. J. O'Toole, H. Abdi, F. Jiang, and P. J. Phillips: "Fusing face-verification algorithms and humans", IEEE Transactions on Systems, Man, and Cybernetics, Part B, 37(5), pp. 1149-1155, 2007.

[23] S. A. Haslam and C. McGarty, Research Methods and Statistics in Psychology, SAGE Publications, 2003.

[24] D. I. Perrett, A. J. Mistlin, and A. J. Chitty, "Visual neurones responsive to faces", Trends in Neurosciences, 10(9), pp. 358-364, 1987.

[25] Y. Yacoob and L. Davis, "Smiling faces are better for face recognition", International Conf. on Face Recognition and Gesture Analysis, Washington DC, pp. 59-64, 2002.

[26] R. S. Malpass, L. A. Zimmerman, C. A. Meissner, S. J. Ross, M. E. Rigoni, L. D. Topp, N. Pruss, C. G. Tredoux, and J. M. Leyva. "Eyewitness memory and identification", The San Antonio Defender, 2005.