

Recognizing Face Sketches by a Large Number of Human Subjects: A Perception-Based Study for Facial Distinctiveness

Yong Zhang¹, Steve Ellyson², Anthony Zone¹,
Priyanka Gangam¹, John Sullins¹, Christine McCullough³
Departments of Computer Science¹, Psychology², Art³
Youngstown State University, Youngstown, OH 44555, USA

Abstract—Understanding how humans recognize face sketches drawn by artists is of significant value to both criminal investigators and researchers in computer vision, face biometrics and cognitive psychology. However, large scale experimental studies of 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 paper, we reported the results of a series of psychological experiments in which 406 volunteers were asked to recognize 250 sketches drawn by 5 different artists. The primary findings are: (i) Sketch quality (artist factor) 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 the 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.

Keywords-sketch; artist; forensic; distinctiveness; psychology;

I. INTRODUCTION

Face sketching has been used in criminal investigations since the late 19th century, and continues to be an important forensic technique for law enforcement agencies today [1, 2]. In many cases when other forensic evidences are not available, constructing a face sketch is the last resort to solve a crime. Despite its success and popularity, face sketching is known to have large uncertainties, which in the worse scenario could lead to the false conviction of innocent people [3].

One source of uncertainties is that, in a typical criminal investigation, sketches were almost always identified by humans. In comparison to algorithm based face recognition where variables can be somewhat controlled, the study of sketch recognition by humans is far more challenging because a person's cognitive process is subjective to a variety of factors. For example, human performance can be affected by the exposure time to a face, the distinctiveness of a face, the number of sketches used in a test, the emotion and motivation of a participant, the diversity of perception ability of a population, and the quality of sketches. Therefore, to gain a thorough understanding of how humans identify a subject from sketches demands psychological experiments of a larger sample set, in terms of both the numbers of sketches and

This work was supported in part by Youngstown State University Research Council Grant No. 08-8, NYSTAR James Investigator Program, and the NSF under grants IIS-0414029 and IIS-1051103 at Binghamton University.

Shaun Canavan, Lijun Yin

Department of Computer Science
State University of New York at Binghamton
Binghamton, NY 13902, USA

participants, upon which statistically meaningful conclusions can be inferred.

In cognitive psychology, multi-choice test has been used to study the uncertainties and indecisiveness in decision making strategies [4, 5]. In sketch recognition tests, when given multiple sketches, volunteers showed their hesitation in the form of multi-match or second-guess. We are particularly interested in three issues: (i) Baseline analysis of the overall performance of all participants; (ii) Analysis of participants who did multi-match, and whether their cognitive behaviors reveal more information about facial distinctiveness; (iii) Analysis of volunteers who did second-guess. Linking those cognitive measures to facial distinctiveness is appealing, because distinctive features could be critical for a correct identification given a short exposure time.

To our knowledge, this is the first work that used a large number of volunteers to recognize hundreds of sketches drawn by different artists. Other than its apparent value to forensic research, this study has strong implications to automatic face recognition. As pointed out in [6], “An understanding of human visual processes involved in face recognition can facilitate and, in turn be facilitated by, better computational models”.

II. RELATED WORKS

Face sketch recognition is primarily studied in two communities: computer vision and cognitive psychology (in connection with law and forensic science). We provide a brief review of several most relevant papers and refer readers to [7, 8, 9, 10, 11, 12] for more comprehensive discussions on face perception and recognition.

In computer vision, the research focus is on developing efficient algorithms that search a sketch in a photograph database. Searching was often facilitated by pre-processing methods that bring sketches closer to photos in a projected space [13]. Tang and Wang [14] developed a more sophisticated transformation function based on the eigenface method that largely enhanced the similarity between sketches and photos. This was a significant step because the intensity disparity between sketches and photos could cause serious problems to algorithms. More recently, they utilized a multi-scale Markov Random Fields model to synthesize a sketch from a photo or *vice versa*, so that sketch-photo matching can be done in a straightforward manner [15]. Another study of using a mug-shot database has shown that both local and global features could improve the searching accuracy [16].

In cognitive psychology, efforts have been made to building computerized face composite systems. Frowd *et al.* [17] integrated holistic features into a system based on the psychological parameters. They compared the performances of commonly used composite systems and found that a well-trained police artist was superior to software kits [18, 19, 20]. Another finding is 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 [21, 22].

III. SKETCH DATASET

A dataset of 250 hand-drawn sketches was used in the experiments. First, frontal-view face images of 50 subjects were randomly selected from a database, all with a neutral expression. The grayscale images were then printed on white papers (photographs). An artist drew a sketch by looking at a photograph. Depending on the drawing style of an artist, a sketch took 30 to 60 minutes to complete. Five artists (a professors and students of art at a university) participated in the project. All artists received special instructions on how to draw forensically relevant sketches in a workshop given by a trained police artist. Each artist drew 50 sketches, one for each of the 50 subjects. Examples of a photo and sketches are shown in Figure 1.

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

Figure 1: A face photo and the corresponding sketches drawn by five artists. The photo and sketches were normalized using the eye coordinates.

IV. EXPERIMENTS AND METHODS

A. Volunteers

A total of 406 volunteers participated in the tests. The volunteers were primarily college students who were taking an introductory psychology course at the time of study. The majority of students were freshmen or sophomores. They represented a wide range of education backgrounds as reflected in their majors, including those in College of Liberal Arts and Social Sciences, College of Science, Technology, Engineering, and Mathematics, College of Education and Business School. Therefore, the volunteers’ performance can be considered representative of a much larger population.

B. Experiment Design and Protocol

The experiments were carried out in three separate sessions. The first session had 61 volunteers, the second session had 184 volunteers, and the third session had 161 volunteers. Prior to each session, a researcher explained the nature and scope of the experiments as well as the risk and benefits of participating in

the study. All participants signed an IRB (Institutional Review Board) approved consent form.

During each session, a volunteer was given two papers: (1) A reference sheet (50 photos in the first session and 25 photos in the second and third sessions, with each photo corresponding to a different person); (2) An answer sheet (50 sketches in the first session and 25 sketches in the second and third sessions, with each sketch corresponding to a different person). The reference and answer sheets contain the photos and sketches of the same 25 or 50 persons, but with their positions randomized. Examples of reference/answer sheets were given in Figure 2. The sketches in an answer sheet were drawn by the same artist. For each sketch, participants were asked to find a match in the photo sheet. Two special cases were considered: multiple-match and second-guess. A sketch can be matched to multiple photos and *vice versa*. If a participant found a better match, he or she should mark the original photo ID on the answer sheet and write down a new ID. This tells us which match was a result of second-guess.

Since the first session had twice number of photos and sketches (50) than the second and third sessions had (25), it took the participants of the first session 30-60 minutes to finish the test, while the participants of the second and third sessions were able to complete the task in less than 20 minutes. To minimize the impact of fatigue and loss of concentration, light refreshments were provided in the first session. Information of the volunteers and test conditions were summarized in Table I.

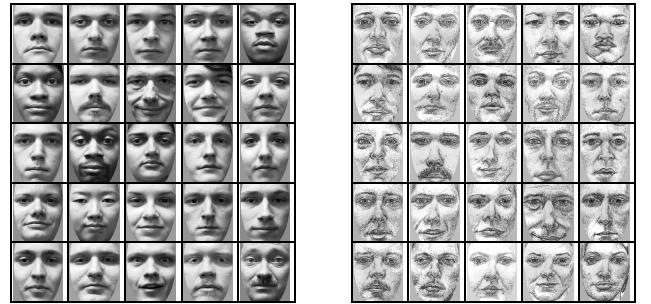


Figure 2: Examples of reference sheet (photos, left) and answer sheet (sketches, right). The sketches were drawn by Artist-5.

TABLE I
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

C. Evaluation Methods

The results of each session were tabulated into a 2D score matrix, with each row for a subject in sketches and each column for a volunteer (Table II). A binary counting method was used: 1 for a hit (correct match), and 0 for a loss (incorrect match). The total number of hits along a row or a column was calculated and used to compute two recognition rates:

$$R_s = H_{row}/N, \quad R_v = H_{col}/M,$$

where H_{row} is the number of hits along a row (the number of correct matches by all volunteers with respect to a sketch), H_{col} is the number of hits along a 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. In other words, R_s represents the recognition rate of a sketch averaged over all volunteers, and R_v represents the recognition rate of a volunteer averaged over all sketches. In this study, R_v was mainly used to analyze the performance variation of volunteers, while R_s was used to characterize the subjects in sketches. For example, which face portrayed in sketches was more recognizable than the others? In the analysis of second-guess, only the cells that correspond to a second-guess mark in the answer sheet were used in the calculation. Similarly, only the cells that correspond to actual multi-matches in the answer sheet were counted for the subsequent multi-match analysis.

Because of the nature of the study, we assessed human performance using statistical inference methods that have been widely accepted in experimental psychology, including ANOVA and post hoc tests. In this way, our results can be compared to that of similar studies.

TABLE II

THE SCORE MATRIX FOR CALCULATING RECOGNITION RATES (R_v, R_s)

	Volun. 1	Volun. 2	...	Volun. 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		

V. RESULTS AND DISCUSSIONS

A. Baseline Performance

Three issues were 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?

The statistics of the recognition rate (R_v) were given in Tables III and IV (the results of the first session were reported separately because its answer sheet had 50 sketches). The mean recognition rates of five artists were plotted in Figure 3. Artist-4 had the highest recognition rates, followed by Artist-2 and Artist-5. As expected, the rates of using 50 sketches were lower than that of using 25 sketches. But the relative performances of five artists remained the same.

Two one-way ANOVAs were conducted, with artist as the categorical factor and R_v as the dependent variable. The first ANOVA used the data of the first session (61 volunteers, 50 sketches). The Levene test suggested no violation of homogeneity of variance assumption across groups: $F(4, 56)=1.93, p=0.12$. The differences among 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 Artist-3.

The second ANOVA used the results of the second and third sessions (345 volunteers, 25 sketches). No violation of homogeneity of variance was found by the Levene test: $F(4, 340)=1.14, p=0.34$. Again, there was a significant variation among artists: $F(4, 340)=28.99, p<0.001$, with an effect size of $\eta^2=0.25$. The pair-wise comparisons of Tukey's test ($\alpha=0.05$) generated similar rankings of artists as in the first ANOVA.

From the above data, it can be concluded that the sketches of Artist-1 are most challenging, while the sketches of Artist-4 are of the highest quality. The fact that Artist-4 was an Art Professor and all other artists were students implies that experience and training probably play an important role in constructing a reliable face sketch.

TABLE III
RECOGNITION STATISTICS OF THE 1ST SESSION
(50 SKETCHES IN THE ANSWER SHEET, 61 VOLUNTEERS)

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 IV
RECOGNITION STATISTICS OF THE 2ND AND 3RD SESSIONS
(25 SKETCHES IN THE ANSWER SHEET, 345 VOLUNTEERS)

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%

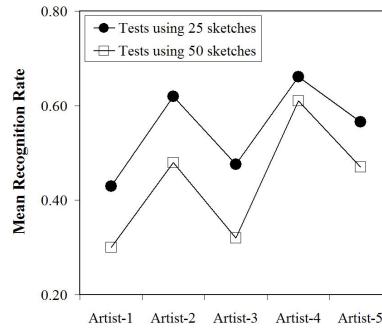


Figure 3: The mean recognition rates (R_v) of the first session (50 sketches) and the second and third sessions (25 sketches).

Choosing an appropriate number of sketches in a test is critical to reveal the true cognitive capability of participants, because it is related to the fatigue effect that could undermine the internal validity of a psychological experiment [23]. As the number of sketches increases, participants' performance could drop quickly due to the loss of concentration and interest. On the other hand, if a very small number of sketches were used, the recognition rates could be so high that suppress the cognitive diversity of participants. In Figure 4, we plotted the R_s values of two test designs (25 sketches versus 50 sketches). The good correlation between the two designs ($r=0.84, p<0.01$)

and their balanced recognition rates (Table III and Table IV) suggest that using 20-50 sketches is a reasonable choice. For a test of a larger scale with a tight time constraint, 10-15 sketches may be considered so that participants can finish the test in less than 10 minutes.

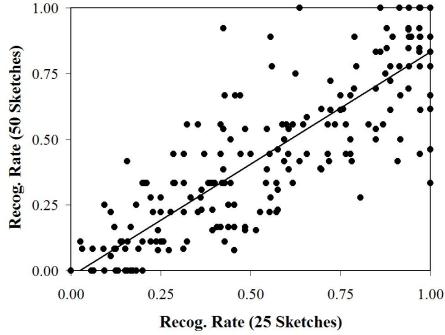


Figure 4: The correlation of recognition rates (R_s) between the test of using 25 sketches and the test of using 50 sketches.

A recent study of computer generated composites found that facial distinctiveness is an important factor in target naming [18]. This is in agreement with our daily life experience as well as neurobiological findings that certain neurons are more responsive to a particular facial part or feature [24]. It is therefore strongly desirable to devise a rating system that can quantify the distinctiveness of a face sketch. Such a system should take into account the performance measures of a large number of human evaluators and multiple algorithms. For this purpose, we ranked all subjects appeared in sketches based on their recognition rates (R_s) averaged over five artists, and showed a few of them in Figure 5 (six from the top ten and six from the bottom ten).

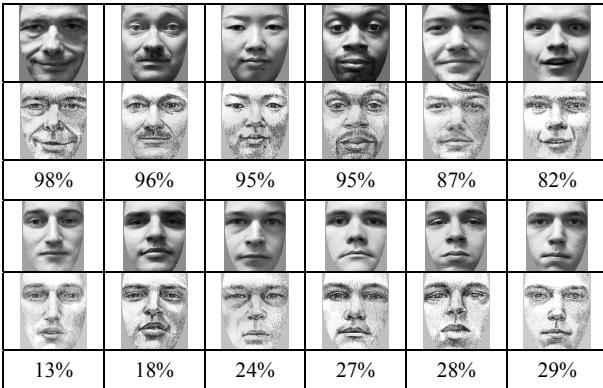


Figure 5: The most and least recognized faces as ranked by their recognition rates (R_s) averaged over five artists. All sketches were drawn by Artist 2.

The highly recognized faces had several characteristics: (i) In older age groups (two subjects > 40 years old, one subject > 60 years old); (ii) Having either mustaches or hair; (iii) Minority ethnic groups (one Asian female, one African American male); (iv) Having facial expressions. It is not surprising that human vision is more responsive to those salient visual stimuli. For example, it has been found that smiling faces are better for recognition [25]. On the other hand, the less recognized faces converged to the “average face” that is typified as young, clean, and of neutral expression. The high

similarity among those faces also lead to the large number of multiple matches to be discussed in the next section.

B. Multi-Match

In a forensic investigation, eyewitnesses or victims may be asked to positively identify a culprit from a group of possible suspects, in the form of either mug-shots or face composites. Studies have shown that the procedures of line-ups and 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]. A similar problem occurred in sketch recognition tests when volunteers found that a sketch looks like multiple photos or *vice versa*. It is essential to know how multi-match affects the human performance. More importantly, multi-match could be a useful measure of facial distinctiveness that can be harnessed in developing better recognition algorithms or face composite systems.

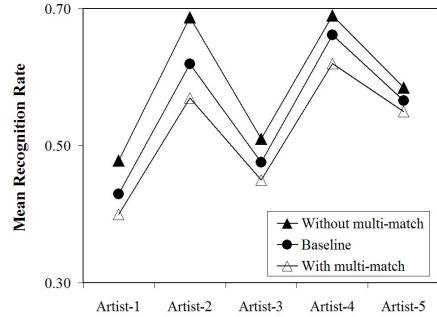


Figure 6: The comparison of volunteers who did multi-match versus those who did not do multi-match, as well as the baseline data. The mean recognition rates were computed using R_s .

Two types of multi-match were allowed: “One sketch to multiple photos” and “One photo to multiple sketches”. Since less than 2% of volunteers did the first type of multi-match, we focused on the second type. The statistics of multi-match were given in Table V. The multi-match ratio is defined as the percentage of volunteers who did multi-match to a particular sketch. The mean recognition rates (R_s) were plotted in Figure 6. As expected, volunteers who did multi-match had lower recognition rates than those who did not. Since the baseline data consists of all volunteers, its performance lies in between. It is interesting to find that only 8% volunteers hesitated on the sketches of Artist-4, much less than that of other artists. This was likely attributed to the higher quality of sketches by Artist-4 as had been demonstrated in the baseline analysis.

TABLE V
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%

We also conducted a two-way ANOVA. The R_v values of 345 volunteers from the second and third sessions were subjected to a 5×2 factorial ANOVA with artist and multi-match as between subject factors. Levene test indicated no violation of 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 in 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 V and Figure 6.

The multi-match ratios averaged over five artists were plotted against the mean recognition rates (R_s) in Figure 7. It appeared that the multi-match ratio was correlated with the recognition rate ($r=-0.79$, $p=0.01$). We also ranked faces based on their multi-match ratios and then showed a few representative ones in Figure 8 (picked from the top ten and the bottom ten). The six faces that received the fewest multi-match hits were also highly recognized in baseline case (Figure 5), with a slightly different order. The faces of higher multi-match ratios also correspond well to the least recognized faces. This suggests that the multi-match ratio could be an important parameter of measuring facial distinctiveness.

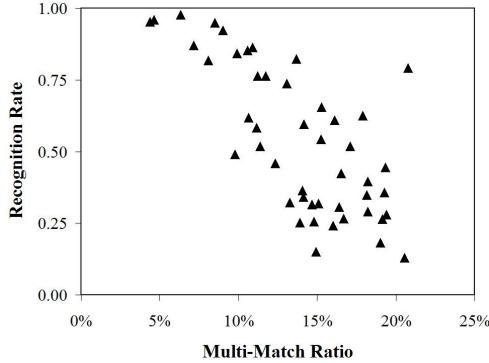


Figure 7: The relationship between the recognition rate (R_s) and the multi-match ratio. The sketches of higher multi-match ratios tended to have lower recognition rates and vice versa.

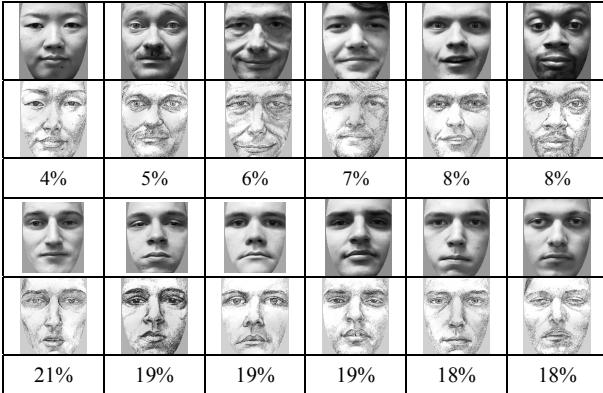


Figure 8: The least and most multi-matched faces as ranked by their multi-match ratios averaged over five artists. All sketches were drawn by Artist 4.

C. Second-Guessing

Second-guessing is an intensively studied topic in social and behavioral psychology. Research evidence showed that second-guessing was actually beneficial to people who changed their test answers (first instinct fallacy), because a complex thinking process was involved [4, 5]. In sketch recognition tests, we want to know whether second-guessing had a positive or negative impact on participants' recognition rates.

The recognition rates of volunteers who did and did not second-guess their initial choices were given in Table VI. The second-guess ratio is defined as the percentage of volunteers who second-guessed a particular sketch. A two-way ANOVA with a 5×2 fully factorial design was carried out. Artist and second-guess served as two factors, and R_v of the second and third sessions (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: $F(4,335)=0.72$, $p=0.58$.

Apparently, second-guessing did not help nor hurt volunteers' overall performance. One possible explanation is that only a small percentage of sketches were second-guessed. The second-guess ratios (Table VI) were much lower than the multi-match ratios (Table V). The lack of correlation between the second-guess ratio and the multi-match ratio also implies that the thinking processes involved in multi-matching and second-guessing are probably independent of each other.

TABLE VI
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%

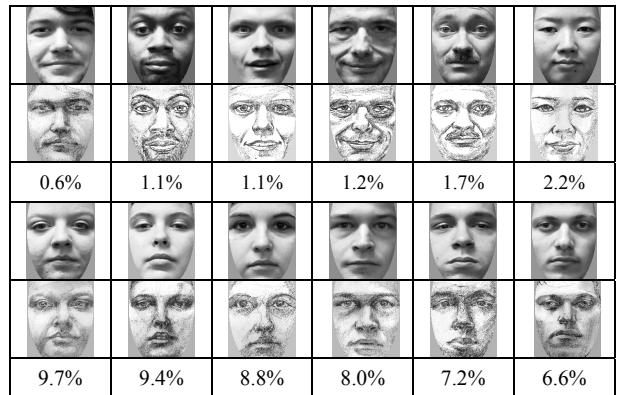


Figure 9: The least and most second-guessed faces as measured by their second-guess ratios averaged over five artists. All sketches were drawn by Artist 3.

Unlike in the multi-match case (Figure 7), no correlation was observed between the second-guess ratio and the recognition rate (R_s). We also ranked subjects based on their second-guess ratios (Figure 9) and found that the least second-guessed faces matched well with the highly recognized faces in baseline analysis (Figure 5). However, the faces that received the most second-guess hits are quite different from the least recognized faces. As a result, the second-guess ratio may still be considered in facial distinctiveness calculation, but probably only for the less second-guessed faces.

VI. SUMMARY

We have presented a sketch recognition study involving a large number of human participants and hundreds of sketches of different drawing styles. Although the analysis was done primarily from a psychological perspective, the findings have strong implications to algorithm based face recognition, biometrics and pattern recognition in general. We summarized the major findings below:

- The sketch quality had a significant effect on human recognition rate. The effect has been observed in all ANOVA results, regardless of the data being analyzed (baseline, multi-match and second-guess). It is apparent that the artist factor (or composite system factor) has to be taken into account when designing a similar sketch recognition test, especially those under a more forensically realistic condition.
- Some faces were consistently ranked as highly recognizable, whether the ranking were based on recognition rate, multi-match ratio or second-guess ratio. The high consistency suggests that those faces possess unique features that can be explored in quantifying facial distinctiveness. The recognition rate, multi-match ratio and second-guess ratio are potentially important parameters to be considered in an objective rating system of facial distinctiveness.
- More sketches were multi-matched than being second-guessed. The multi-match ratio appeared to be correlated with the recognition rate, whereas second-guessing did not alter human performance significantly.
- Test results of using 25 sketches and 50 sketches were consistent, suggesting that 25-50 is a reasonable number of sketches to be used in future investigations with a similar experimental protocol.

Our future work will also study the individual facial distinctiveness and their quantitative measurements using 3D sketch model morphing and topographic feature distributions in order to quantify facial distinctiveness to a set of facial model parameters. In addition, we will also investigate the facial distinctiveness from facial dynamics with various facial expressions so that a better computer-based recognition algorithm can be designed.

REFERENCES

- [1] K. T. Taylor, *Forensic Art and Illustration*, CRC Press, 2000.
- [2] D. McQuiston-Surrett, L. Topp, and R. Malpass, "Use of facial composite systems in US law enforcement agencies", *Psychology, Crime and Law*, 12(5), pp. 505-517, 2006.
- [3] G. Wells and L. Hasel, "Facial composite production by eyewitnesses", *Current Directions in Psychological Science*, 16(1), pp. 6-10, 2007.
- [4] J. Kruger, D. Wirtz, and D. Miller, "Counterfactual thinking and the first instinct fallacy", *Journal of Personality and Social Psychology*, 88(5): 725-735, 2005.
- [5] L. Benjamin, T. Cavell, and W. 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. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, 35(4):399-458, 2003.
- [9] S. Li and A. Jain (editors), *Handbook of Face Recognition*, Springer, 2005.
- [10] K. Bowyer, K. Chang, and P. 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 IEEE CVPR 1996*, pp. 586-593.
- [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 Trans. on PAMI*, 31(11):1955-1967, 2009.
- [16] P. Yuen, and C. Man, "Human face image searching system using sketches", *IEEE Trans. on SMC-Part A*, 37(4):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. Frowd, D. Carson, H. Ness, J. Richardson, L. Morrison, S. McLanaghan, & P. Hancock, "A forensically valid comparison of facial composite systems", *Psychology, Crime & Law*, 11(1):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 PAMI*, 29(9):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. Perrett, A. Mistlin, and A. 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. Malpass, L. Zimmerman, C. Meissner, S. Ross, M. Rigoni, L. Topp, N. Pruss, C. Tredoux, and J. Leyva. "Eyewitness memory and identification", *The San Antonio Defender*, 2005.