Evidence on Skill Differences of Women and Men Concerning Face Recognition

Josef Bigun, Kwok-wai Choy, and Henrik Olsson

Halmstad University, Box 823, S-301 18 Halmstad, Sweden

Abstract. We present a cognitive study regarding face recognition skills of women and men. The results reveal that there are in the average sizable skill differences between women and men in human face recognition. The women had higher correct answer frequencies then men in all face recognition questions they answered. In difficult questions, those which had fewer correct answers than other questions, the performance of the best skilled women were remarkably higher than the best skilled men. The lack of caricature type information (high spatial frequencies) hampers the recognition task significantly more than the lack of silhouette and shading (low spatial frequencies) information, according to our findings. Furthermore, the results confirmed the previous findings that hair style and facial expressions degrades the face recognition performance of humans significantly. The reported results concern 1838 individuals and the study was effectuated by means of Internet.

1 Introduction

The seminal study of Bruce and Young [2] aims to develop a theoretical model and a set of terms for understanding and discussing how we recognize familiar faces, and the relationship between recognition and other aspects of face processing. Faw [6], who, along with Bruce and Young [3], we refer to for an extensive review on human skills of face recognition, compares 112 human studies about face recognition and investigates the need for a standardization.

Hassing et. al. [7] studied the relative importance of age, gender and education on episodic memory functioning in a population-based sample of healthy individuals, between 90 and 100 years of age. The results, that do not quantify the male and female skills, suggest that education, age and gender has no effect in face recognition skills in this age category. Other age categories were not studied. Twenty test pictures were shown to subjects during 6 seconds to subjects who should later recognize these among 20 distractor pictures. The observed face recognition skills of the subjects were modulated by the memory skills.

Bruce and Beard [1], studied female African American and Caucasian Americans' ability to recognize female faces of their own racial group and/or another racial group. They report on two experiments. In the Experiment 1, participant subjects saw either African American or Caucasian American faces; in Experiment 2, all participants saw faces of both races. There was no evidence of cross-racial bias in Experiment 1. Experiment 2 revealed some evidence of cross-racial

J. Bigun and F. Smeraldi (Eds.): AVBPA 2001, LNCS 2091, pp. 44–50, 2001.

bias, in which Caucasian Americans performed more poorly and made more errors in recognition of African American faces than the African Americans did on Caucasian Americans.

The Priming effects in children's face recognition are studied by Ellis et.al. [5]. The subjects (children and young adults) were told to describe either face expression or gender. Subsequently some familiar person's pictures among unfamiliar persons' pictures were shown for a judgment. There were three age categories (five-, eight-, eleven-year old). One half were assigned to judge expression and the other half to judge gender. The reaction time was estimated with the computer in order to see how fast or good the face recognition was. The experiment indicates that participants of five years show the same reaction time as older children in face recognition. In this experiment the pictures were the same on each occasion. In the second experiment they also showed a different view of the face and the result concerning the reaction time was the same. Consequently, there are not developmental changes in the implicit memory. According to this study there are developmental changes in the explicit memory.

We present below original evidence that uncovers the significant skill differences between the genders as they apply to face recognition based on a large cognitive study. We also present evidence confirming known distractor factors presented in other studies.

2 Method

The experiments were conducted in a spatially distributed manner roughly during one month by using the Internet and web pages.

2.1 The Host Program

Upon starting the test, a Java program, which we will here refer to as the *host program,* specifically written for the present study, walked the subject through the various web pages, step by step until all answers were registered. The host program had a number of questions attempting to quantify some face recognition skills that will be described in detail further below. In the design of the host program, special attention was given to that the subjects were prevented to give more than one answer to a question and that once a question was answered they were only allowed to go to the next question by pushing the "Next" button. Consequently, the questions had to be answered one after the other but the subject decided without constraint on time when to fill the answer and thereby move on to the next question.

The following compulsory information was gathered from every subject who completed the test:

1. *recognition data* consisting of the answers of the subject to the face recognition questions,
2. *subject data* consisting of the physical (gender and age) and geographical data (city type and continent) of the subject.

Fig. 1. The test object (left-top) of Question 1, to be identified among the 10 objects underneath. On the right, the analogous data for Question 2 are given.

Fig. 2. The test object (left-top) of Question 3, to be identified among the 10 objects underneath. On the right, the analogous data for Question 4 are given.

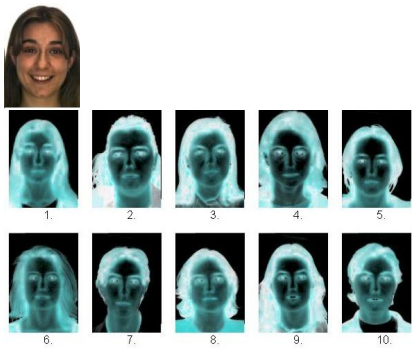

Table 1: Aggrageted correct answers

Ex: Column 7 says 14 % of F:s and 5 % of M:s had 7 correct answers.

	0	1	2	3	4	5	6	7	8	Total
Fem.	0%	2%	3%	13%	20%	23%	24%	14%	2%	100%
Male	2%	6%	14%	20%	21%	20%	12%	5%	0%	100%
Fem.	0	4	8	35	52	61	62	37	4	263
Male	11	44	97	133	144	135	79	34	3	680

Table 2: Question and gender wise correct answers

Ex: Column Q2 says 16 % of F:s and 6 % of M:s had correct answers to Q2.

F: 263 M: 680	Edges Q1	Blur Q2	M-F-P Q3	F-F-P Q4	MemM Q5	MemF Q6	Eyes Q7	Negs. Q8
Fem.	48%	16%	82%	70%	76%	82%	44%	79%
Male	31%	6%	72%	47%	68%	66%	33%	63%
Fem.	125	42	217	183	201	215	114	208
Male	209	39	490	322	465	447	222	431

Fig. 3. The images used in Question 8. Tables 1 and 2 illustrate the aggregated CAs and detailed answer statistics.

Terminating the host program by closing the window was of course possible but no registration of the data took place at Halmstad University server in that case. In addition to the compulsory information the following voluntary information was gathered from the subjects:

– name, address, and E-mail.

2.2 The Face Recognition Questions

The test consisted of 8 questions, (Q1,...Q8). The task was to identify the picture of an object person among a set of 10 pictures of persons, object set. In Q1-Q4, and Q8 the test object was shown on the top of the object set simultaneously, in the same page. The Q5-Q7 were similar to other questions except that they included a memory task i.e. the test object was shown in its own page, after the close of which by the test subject, a page containing only the object set was presented.

The subjects were informed before the start of the test object image and the image to be found in the object set were taken at two different occasions, i.e. they were told that, the two images could differ significantly in hair style, glasses, expression of the face, facial hair, clothing, ... etc due to the natural changes in appearance that occur upon passage of time (a few months). The images were 100x144 pixels each, except the test object of Q7, representing an eye region image of size 180x67 pixels. All test objects were displaying either full frontal or full profile views of the head, except in Q7. Except in Q8, in which the images contained 256 colors, all images lacked color but had 256 gray tones.

Q1 (Edges): The test object is a caricature image of a man, Figure 1. This image is obtained from a real photograph of the test object by a spatial high pass filtering. This filtering highlights certain edges by suppressing low spatial frequencies. Here it is achieved by taking the difference of the highest two resolutions (scales) of a Gaussian pyramid, obtained by using a true Gaussian filter with $\sigma = 1.22$, [4].

Q2 (Blur): The test object is a smoothed image of a woman, Figure 1. In contrast to question 1, the 10 test objects contain now only low frequencies. The low frequency images were obtained by using a Gaussian filter with $\sigma = 4.67$.

Q3 (Male-F-P), and Q4 (Female-F-P): The faces of the tests objects have frontal views whereas the faces of the object sets have profile views, Figure 2.

Q5 (MemM), Q6 (MemF): Face images are similar to Q3 and Q4 respectively except that the memorization task was added.

Q7 (Eyes): The test object consisted of the frontal eye region cut out of a photograph of an adult man. The picture did not show nose, mouth, or hair line. The object set displayed full frontal views of the face as in the object set of Q1. The question had the memorization task.

Q8 (Negatives): The test object was a color image of a female, whereas the images of the objects set were brightness and hue negated so that they looked like negative photographs, Figure 3 left. The negation was done in the HSV color space.

3 Experimental Results

Approximately 10'000 persons, with interest in science and technology in areas related to pattern recognition, were invited to do the test via www. As of the time of writing, 1 month after the invitation, 1838 persons, 492 women and 1346 men, took the test. Male dominance is a consequence of the male dominance in the call list. The fact that it is trivially simple to send false information on gender, age, ..etc has been handled as a statistical noise in our case for two reasons: i) we invited only a certain category of people, those who have related professional interests or activities to take the test, ii) nearly half of the female and male subjects did also send their names, addresses and emails. This enabled us to use the non-anonymous group as a control group for the anonymous group. Based on this we conclude that the noise from false data was negligible.

Given the number of categories (8 age categories, 8 continent categories, 2 gender categories), the number of participants, and the scope of this paper, we only report on the age category 21-30 and all continents. This age category contained 263 women and 680 men in our study. The statistics can be assumed to be representative for Caucasians since approximately 90 % of the participants were from Europe and USA. Although we report on one age category (the one on which we had most data), *also other age categories with significant number of subjects, confirm the conclusions reported below.*

The correct answers (CA) distribution represents the portion of the subjects that gave 0, 1, ...8 correct answers to the 8 questions across the male and the female subjects, Figure 3. The CA distribution evidences that the female subjects were more skilled to give the correct answers than the male subjects. The bias was observable even if the continent and the town type was restricted to be a small town (Halmstad, Sweden, 65'000 population) in Europe. The more correct answers, the less likely that these came from the male population e.g. it was approximately 2 times more likely that it was a woman who gave 6 correct answers than a man, 3 times more likely that it was a woman who gave 7 correct answers than a man...etc In other words the best female skills were far better than the best male skills although the female subject population had even in the average (median) more correct answers (5) than the male population (4).

Another novel result is that the high spatial frequencies carry more significant information for human face recognition than low spatial frequency information. This is evidenced by the fact that Q2 was much more difficult to answer correctly than Q1 (as well as all other questions), See Table 2. A totally random answer to Q2 would result in 10 % success which is suspiciously close to the female (16 %) and male (6 %) success rates. However the result of the subjects who have succeeded in this difficult question appears to be more due to skills than "luck". This is even more so for females than males because the females who had succeeded in Q2 were the top performers in all questions: they had in the average 6.2 (M: 5.4) correct answers with variance 1.5 (M: 2.5). Unskilled female and male subjects would have been represented equally well among those succeeded in Q2, had the results been due to random effects. In terms of the occurance of

the next maximum in votes of Q2, both genders' agreed on the same answer: the correct answer. This is another support for skill as explanatory factor.

Our finding on the importance of high frequency data for human face recognition should be contrasted to the unimportance of these for some machine face recognition techniques based on small face pictures, i.e. lack of high frequencies, due to computational considerations.

For both women and men, the hair style was a significant distractor which is a reconfirmation of the results from previous studies, e.g. Bruce and Young, [3]. This was mostly evident from the answers to Q1, Q2, and Q8. The largest incorrect answers were always due to distractors had hairlines similar to hairlines of the test objects. This phenomenon was particularly striking in the CAs of Q2, in which 51 % of the female and 63 % of the male subjects were convinced that an erroneous answer was the correct answer. For both genders the largest (erroneous) answer coincided and the frequencies for the correct answers (F: 16 % M: 6 %) were the second largest. The hair style of the corresponding distractor that could cause so many subject's failure, was similar to that of the original test object, whereas the test object had changed her hair style completely in the image to be matched.

Q4 is a matching task between a frontal and a profile view of female objects, CA frequency were F: 70 % M: 47 % for male subjects. For both genders, Q4 was obviously more difficult to answer than Q3, having the CA frequencies F: 82 % M: 72 %, which concerns matching of a male object's frontal and profile views. The additional difficulty in Q4, resided in that there were two distraction elements: the hair style and the facial expression. From previous studies it is known that the recognition is hampered by a change in facial expression or hair style. But that the drop in performance is more significant for men than women has not been reported previously. By contrast, it is not easy to cross compare the CA's of Q3 and Q4 due to unequal difficulty level of the questions. In other words the numerical closeness of the CA's 72 % (female subjects on female objects, in Q4) and 70 % (male subjects on male objects,in Q3) concern questions of two different scales of difficulties.

With respect to Q5 and Q6, results show once more that the females performed significantly better. Again cross comparison of CA's between questions is more speculative than a comparison within the questions due to the different difficulty scales in the questions. For example one would be tempted to conclude that males recognizing females improved significantly in Q5 and Q6 as compared to questions Q3 and Q4. By contrast, the closeness of the CAs of males in Q5 and Q6 suggests that Q3, Q5 and Q6 were nearly equally difficult to answer. Independently, we find indeed that the female CA's for the same questions are nearly the same (82 %, 76 % and 82 %) too. This suggests that there is no additional bias of the memory on the skills of the genders, since the CA performance difference between the genders were nearly not altered by the memory capabilities.

The performance of the subjects in Q7 dropped indeed because the eye information is more restrictive, but the results are significantly above a 10% random selection, confirming that the females performed better than males.

Finally, concerning question 8, females performed again better than males. Apparently, negative images, although looking very different, still provide most if not all features necessary for recognition.

4 Conclusions

We have designed and implemented a human face recognition study that aimed at quantifying the skills of various categories of humans, including gender and age. The results uncovered that the female population have better skills in human face recognition tasks than the male population. We could reconfirm that the hair style and facial expressions are significant distraction factors for humans when they recognize faces. However, a novel finding is that men appear to have larger negative bias caused by these distractions than women.

We also found that the lack of details (high spatial frequencies) hampers the recognition significantly more than the lack of silhouette and shading (low spatial frequencies) information in general.

A more detailed quantification of our findings needs further studies.

Acknowledgment

We gratefully acknowledge the support of the Chalmers foundation and the use of M2VTS face database in our study, [8]. We thank Prof. J. M. H. du Buf, Univ. of Algarve, for his valuable comments on the manuscript.

References

1. A. J. Bruce and K. W. Beard. African Americans, and Caucasian Americans recognition and likability responses to African American and Caucasian American faces. *Journal of General Psychology*, 124:143–156, Apr 1997.
2. V. Bruce and A. Young. Understanding face recognition. *British Journal of Psychology*, 77:305–327, 1986.
3. V. Bruce and A. Young. *In the eye of the beholder*. Oxford University Press, Oxford, 1998.
4. P. Burt. Fast filter transforms for image processing. *Computer graphics and image processing*, 16:20–51, 1981.
5. H. D. Ellis, D. M. Ellis, and J. A. Hosie. Priming effects in childrens face recognition. *British Journal of Psychology*, 84:101–110, 1993.
6. H. W. Faw. Recognition of unfamiliar faces: Procedural and methodological considerations. *British Journal of Psychology*, 83:25–37, 1992.
7. L. Hassing, A. Wahling, and L. Backman. Minimal influence of age, education, and gender on episodic memory functioning in very old age: a population-based study of nonagenarians. *Archives of Gerontology and Geriatrics*, 27:75–87, 1998.
8. S. Pigeon and L. Vandendorpe. The M2VTS multi modal face database (release 1.0). In J. Bigun, G. Chollet, and G. Borgefors, editors, *Audio and Video based Person Authentication - AVBPA97*, pages 403–409. Springer, 1997.