Subjective Experiments on Gender and Ethnicity Recognition from Different Face Representations

Yuxiao Hu[1], Yun Fu[2], Usman Tariq[3], and Thomas S. Huang[3]

[1] Microsoft Corporation, 1 Microsoft Way, Redmond, WA 98052 USA
[2] Department of CSE, University at Buffalo (SUNY), NY 14260, USA
[3] Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA
Yuxiao.Hu@microsoft.com, raymondyunfu@gmail.com,
{utariq2,huang}@ifp.uiuc.edu

Abstract. The design of image-based soft-biometrics systems highly depends on the human factor analysis. How well can human do in gender/ethnicity recognition by looking at faces in different representations? How does human recognize gender/ethnicity? What factors affect the accuracy of gender/ethnicity recognition? The answers of these questions may inspire our design of computer-based automatic gender/ethnicity recognition algorithms. In this work, several subjective experiments are conducted to test the capability of human in gender/ethnicity recognition on different face representations, including 1D face silhouette, 2D face images and 3D face models. Our experimental results provide baselines and interesting inspirations for designing computer-based face gender/ethnicity recognition algorithms.

Keywords: Face analysis, gender recognition, ethnicity recognition, subjective experiment.

1 Introduction

Humans are quite accurate when deciding the gender of a face even when cues from makeup, hairstyle and facial hair are minimized [1]. Bruce et al. showed that cues from features such as eyebrows and skin texture play an important role in decision making as the subjects were less accurate when asked to judge a face based solely upon its 3D shape representation compared to when 2D face images (with hair concealed and eyes closed) were present. According to [1], the average male face differs from an average female face in 3D shape by having a more protuberant nose/brow and more prominent chin/jaw.

The results from [2] shows that both color and shape are vital for humans in deciding the gender and ethnicity from a 2D face image. Color was more important factor in gender decisions while shape was more important in deciding ethnicity. However, when the both sources of information were combined, the dominant source depended on viewpoint. Shape features proved to be more important in angled views while the color in full-face view.

S. Boll et al. (Eds.): MMM 2010, LNCS 5916, pp. 66–75, 2010.

Males, females and different ethnicities also differ in silhouetted face profile, which can be represented in 1D stimulate. For instance, the ridge of the bone above the eye is more pronounced in males [3]. Also the forehead in case of males is backward sloping while that of the females tends to be more vertical. The female nose tends to be more concave and straighter. The distance between top lip and base of the nose is usually longer in males. The chins of males are much taller than in females.

Overall, humans are able to determine the facial categorization by 3D, 2D and 1D cues or their combination. Different to the case of humans, gender and ethnicity recognition present two of the major challenges of human face analysis in human computer interface researches. Gender recognition is useful in face recognition as it can reduce the problem to matching the face with almost half of the database (if both the genders have equal probability of occurrence in the database). Ethnicity identification will reduce the problem even further. Gender and Ethnicity recognition can also be useful in Human Computer Interface (HCI). The computer shall adapt to the person's gender and ethnic group in terms of speech recognition or offering the person options which is more specific and useful to a particular gender or ethnicity.

The earliest technique for gender classification was based upon neural networks [4]. SEXNET, a fully connected two-layer network, was incorporated in [5] to identify gender from face. A PCA based representation along with radial basis functions was employed by Abdi et al. [6]. A mixed approach based upon collection of neural networks and decision trees was employed by Gutta et al. [7]. In [8], a multi-layered neural network is used to identify gender from multi-resolution face images. Better performance has also been claimed by O' Toole et al. [9] using PCA and neural networks. In [10], Jain et al. used Independent Component Analysis (ICA) to represent each frontal face image as a low-dimensional feature vector for gender identification. They reported superior performance of their approach using support vector machine in ICA space with a 96% accuracy rate. In [11], a skin-color based approach has been employed for gender identification.

2 Setup of Subjective Experiments

2.1 Test Face Samples

The face data used in the subjective experiments were collected by our group with Cyberware 3D laser scanner [12]. We randomly selected 100 3D face data samples with balanced ethnicities. Some examples of the face representations are shown in Fig. 1. The gender and ethnic groups' distribution of the face data are shown in Fig. 2.

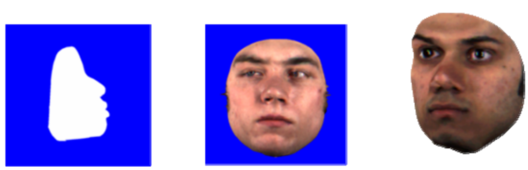

Fig. 1. Different face representations of different faces: silhouette, 2D image, and 3D model

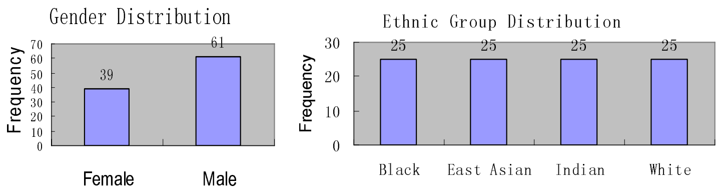

Fig. 2. Gender and ethnics group distribution of the test face

These 100 face data were randomly divided to 3 groups and each group was shown either in silhouette, 2D Face Images or 3D face models representation to each subject. The partition of the data is shown in Table 1.

Table 1. The partition of the test face data

Sample Groups	Number of Samples
SaG#1	33 (#1~#33)
SaG#2	33 (#34~#66)
SaG#3	34 (#67~#100)
Total:	100

2.2 Subjects

Twenty-one graduate students attended the experiments as subjects. They belonged to either “East and South East Asian” or “Indian” or “White” ethnic category. We randomly divided these subjects to 3 groups as shown in Table 2.

Table 2. The partition of the subjects

Subject Groups	Number of Subjects
SuG#1	8
SuG#2	6
SuG#3	7
Total:	21

2.3 Experiments

In the subjective experiments, each subject group was shown three different sample groups in different face representations as shown in Table 3. In this manner, each subject labeled 100 faces, which were grouped into three groups and shown in profile silhouette, 2D Frontal Image or 3D respectively. Each face was shown in three different representations to different subjects, but different representations of a face did not appear in the same test set for a subject. Each face representation was labeled by at least six subjects.

Table 3. The task assignment of different subject groups

Subject Groups	Silhouette	2D texture	3D
SuG#1	SaG#3	SaG#2	SaG#1
SuG#2	SaG#1	SaG#3	SaG#2
SuG#3	SaG#2	SaG#1	SaG#3

Each subject was requested to provide his/her gender and ethnic group information before the experiment. For each face shown on the screen, every subject was asked to judge its gender and ethnicity and specify his/her confidence level of the answer. For gender, the answers could be male or female. For ethnicity, the answers could be one of the following: Black (African or African American), East and Southeast Asian (Chinese, Japanese, Korean, Vietnamese, Filipino, Singaporean ... American Indian), South Asian (Indian, Pakistani ...), White (Caucasian, Hispanic/Latino, Middle Eastern). And the confidence level of above answers could be "Purely guess", "Not sure", "Kind of sure" or "Pretty sure". A screen shot of the experiment program interface is show in Fig. 3, where a profile face is shown.

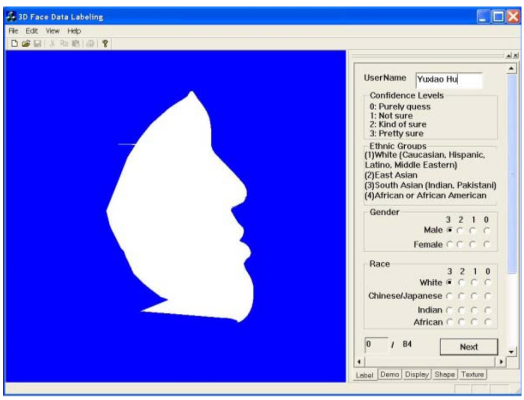

Fig. 3. The experiment program interface

Table 4. Overall accuracy of gender/ethnicity recognition

Accuracy(%)	Gender		Ethnicity	
	Mean	StdDev	Mean	StdDev
Silhouette	57.63%	9.43%	45.08%	18.75%
2D Image	89.41%	6.14%	80.40%	13.52%
3D	92.29%	4.93%	86.00%	12.85%

3 Experimental Results

3.1 Performance Comparison on Different Face Representations

Most of the subjects finished labeling the 100 face samples in 20 minutes. The median time they spent was 896 seconds. Here we didn't use average time because average number is sensitive to outliers. The accuracy of gender and ethnicity recognition was calculated as the ratio between the number of correctly recognized face samples and the total sample number. The overall accuracies of gender recognition on different face representations are shown in Table 4.

It can be seen from these experimental results that gender recognition by human based on silhouette is only a little bit better than chance, which is lower than the accuracy claimed in [13]. On the other hand, ethnicity recognition based on silhouette is significantly above chance. Based upon 2D Image and 3D information, both gender and ethnicity recognitions are pretty accurate. The confusion matrices on gender recognition of different face representation are shown in Table 5. We can see that male faces are better recognized than female faces. As [13] suggested, the reason is that male faces have larger variances on shape and appearances.

The confusion matrices on ethnicity recognition are shown in Table 6. From the confusion matrices on ethnicity, we can see that: Based on silhouette (1D shape), Indians are similar to White and East Asian people; Based on 2D image (texture), Black and Indian are not well separated; Based on 3D information (shape and texture), Black and Indian are better recognized.

Table 5. Confusion matrices on gender recognition based on different face representations

		Recognized As	
Profile Silhouette		Male	Female
GroundTruth	Male	**63.73%**	36.27%
	Female	52.03%	**47.97%**
2D Frontal Face Image		Male	Female
GroundTruth	Male	**94.78%**	5.22%
	Female	18.77%	**81.23%**
3DFace(Shape+Texture)		Male	Female
GroundTruth	Male	**95.33%**	4.67%
	Female	12.55%	**87.45%**

The confusion matrices of the ethnicity recognition accuracy with respect to the subjects from different ethnic group are compared in Table 7. It can be clearly seen that subjects from different ethnic groups recognize the faces of their own ethnic group more accurately.

Table 6. Confusion Matrices on ethnicity recognition based on different face representations

		Recognized As			
Profile Silhouette		White	EastAsian	Indian	Black
Ground Truth	White	64.31%	17.56%	13.47%	4.67%
	EastAsian	25.13%	48.55%	11.71%	14.61%
	Indian	**47.97%**	**26.26%**	**17.67%**	8.09%
	Black	8.75%	14.39%	21.87%	54.99%
2D Frontal Face Image					
Ground Truth	White	83.14%	8.70%	8.17%	0.00%
	EastAsian	5.04%	92.74%	1.68%	0.55%
	Indian	13.24%	5.97%	**60.28%**	**20.51%**
	Black	1.10%	2.72%	**14.05%**	**82.14%**
3D Shape + Texture					
Ground Truth	White	90.58%	7.07%	2.34%	0.00%
	EastAsian	3.45%	95.38%	1.16%	0.00%
	Indian	14.86%	2.28%	**75.43%**	7.43%
	Black	1.13%	2.21%	13.32%	**83.34%**

Table 7. Confusion Matrix of different subject ethnic group based on 3D face representation

		Recognized As			
Indian subjects:		White	EastAsian	Indian	Black
Groud Truth	White	100.00%	0.00%	0.00%	0.00%
	EastAsian	5.57%	94.43%	0.00%	0.00%
	Indian	0.00%	0.00%	**100.00%**	0.00%
	Black	0.00%	0.00%	3.34%	96.66%
Chinese subjects:					
Groud Truth	White	89.31%	8.03%	2.66%	0.00%
	EastAsian	1.74%	**97.41%**	0.85%	0.00%
	Indian	15.33%	2.41%	72.57%	9.70%
	Black	0.85%	2.59%	18.28%	78.27%
Western subjects:					
Groud Truth	White	**90.46%**	7.16%	2.39%	0.00%
	EastAsian	7.33%	90.23%	2.44%	0.00%
	Indian	14.28%	2.04%	81.63%	2.04%
	Black	2.86%	2.86%	5.73%	88.55%

3.2 Performance Comparison of Different Subjects

Among the 21 subjects, the best overall performance of gender and ethnicity recognition is 84%. And the average recognition accuracies of gender and ethnicity recognition are 79.76% and 70.48%. More detailed information about the performance of different subjects is shown in Table 8, Fig. 4, Fig. 5 and Fig. 6; from which we can see that there is no strong correlation between the ethnic group of the subjects and their overall recognition performances.

Table 8. Top ranked subjects on gender and ethnicity recognition accuracy

Rank	Subject No.	Accuracy		
		Gender	Ethnicity	Overall
1	#7, East Asia	85.00%	83.00%	84.00%
2	#9, White	83.00%	78.00%	80.50%
3	#21, Chinese	79.00%	80.00%	79.50%
4	#15, White	83.00%	75.00%	79.00%
5	#6, Indian	86.00%	71.00%	78.50%
Mean Accuracy of All Subjects		79.76%	70.48%	75.12%
Std. Dev. of All Subjects		3.95%	8.45%	4.72%
Max Accuracy of All the 21Subjects		86.00%	83.00%	84.00%
Min Accuracy of All the 21Subjects		73.00%	47.00%	63.00%

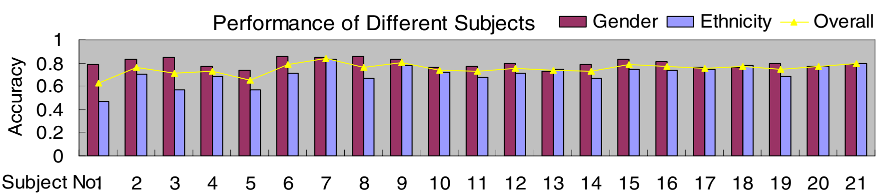

Fig. 4. Performance of gender and ethnicity recognition of different subjects

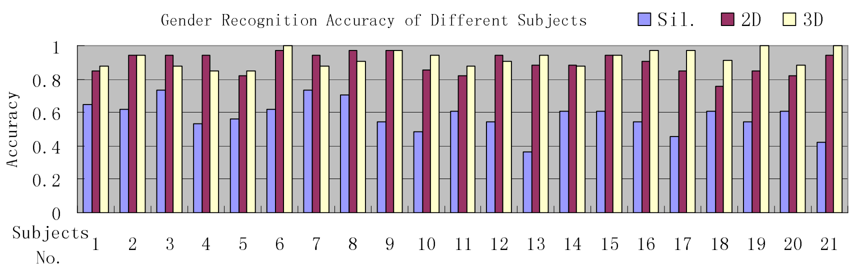

Fig. 5. Gender recognition accuracy of different subjects on different face representations

3.3 Improving the Performance by Combining Multiple Results

In the end, we tried to combine the recognition results of all the subjects by majority vote to see whether this would improve the accuracy of gender and ethnicity recognition. Here each face sample had at least 6 votes in one representation. If there is a tie,

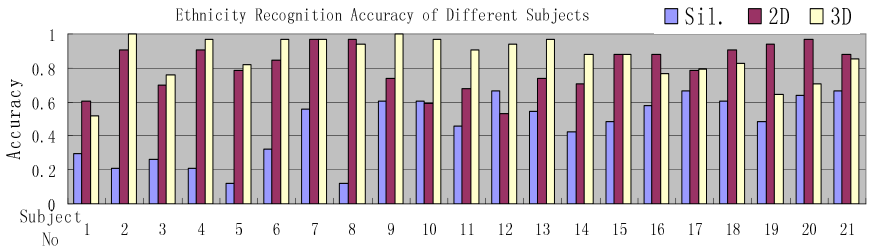

Fig. 6. Ethnicity recognition accuracy of different subjects on different face representations

we would randomly select a result as the decision. We could either vote on the same representation or vote across different representations. The results of voting on same representation are shown in Table 9. The results of voting across different representations are shown in Table 10.

Table 9. Combining the recognition results

Accuracy(%)	Gender		Ethnicity	
	Average	Vote	Average	Vote
Silhouette	57.63%	58%	45.08%	55%
2D Image	89.41%	96%	80.40%	83%
3D	92.29%	97%	86.00%	93%

Table 10. Combining the recognition results across different representations

Accuracy(%)	Gender	Ethnicity
sil+2D+3D	94%	90%
sil+2D	91%	83%
sil+3D	91%	89%
2D+3D	95%	91%

From above results, we can observe that, 1) Voting on gender based on silhouette doesn't help much. 2) Voting on ethnicity based on silhouette increases the accuracy about 10%; 3) Voting on gender/ethnicity based on 2D/3D also improves the accuracy; 4) Combining Silhouette, 2D or 3D does not improve the accuracy. These observations inspire us that if we design multiple classifiers based on different facial feature/representations, their result combination will be better than single classifier on single type of feature.

4 Conclusions

In this work, several subjective experiments were conducted to test the capability of human in gender/ethnicity recognition on different face representations. Based on these results, following conclusions can be made: 1) Gender recognition based on silhouette is slightly better than chance, which indicates the difficulty of silhouette based gender recognition algorithms; 2) Ethnicity recognition based on silhouette is well above chance; 3) Gender/Ethnicity recognition base on 2D/3D information are very accurate. The more information we have, the better the accuracy is; 4) Male faces are better recognized than female faces; 5) Indian faces are better recognized based on 3D information (in balanced ethnic distribution); 6) Subjects from different ethnic groups recognize the faces more accurately from their own ethnic group.

Although above conclusions are drawn from subjective experiments, i.e., the gender/ethnicity recognition is performed by human subjects; computer-based automatic gender/ethnicity recognition algorithms can benefit from these conclusions to infer the effective face representations and discriminant features. That is, computers can "learn from human". Some very encouraging results have already been achieved by our group on silhouetted face profiles. The features used in these experiments were primarily shape context features [14]. In the near future, our plan is to do experiments on other views and come up with an improved classifier for automatic gender and ethnicity identification.

References

1. Bruce, V., Burton, A.M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., Linney, A.: Sex discrimination: how do we tell the difference between male and female faces? Perception 22, 131–152 (1993)
2. Hill, H., Bruce, V., Akamatsu, S.: Perceiving the sex and race of faces: The role of shape and color. Proc. of the Royal Society - Biological Sciences (Series B) 261, 367–373 (1995)
3. The main differences between male and female faces, `http://www.virtualffs.co.uk/male.femalefacialdifferences.htm`
4. Cottrell, G., Metcalfe, J.: EMPATH: Face, Gender and Emotion Recognition using Holons. In: Proceedings of Advances in Neural Information Processing Systems, pp. 564–571 (1990)
5. Golomb, B.A., Lawrence, D.T., Sejnowski, T.J.: Sexnet: a neural network identifies sex from human faces. In: Lipmann, R.P., Moody, J.E., Touretzky, D.S. (eds.) Proc. of NIPS, vol. 3, pp. 572–577. Morgan Kaufmann, San Mateo (1990)
6. Abdi, H., Valentin, D., Edelman, B., O'Toole, A.: More about the difference between men and women: evidence from linear neural networks and the principal component approach. Perception 24, 539–562 (1995)
7. Gutta, S., Weschler, H., Phillips, P.J.: Gender and Ethnic Classification of Human Faces using Hybrid Classifiers. In: Proc. of IEEE Conf. on AFGR, pp. 194–199 (1998)
8. Tamura, S.H., Kawai, M.H.: Male/Female Identification from 8 x 6 Very Low Resolution Face Images by Neural Network. Pattern Recognition 29, 331–335 (1996)
9. O'Toole, A., Abdi, H., Deffenbacher, K., Valentin, D.: A low-dimensional representation of faces in higher dimensions of space. J. of Optical Society of America 10, 405–411 (1993)
10. Jain, A., Huang, J., Fang, S.: Gender identification using frontal facial images. In: Proc. of IEEE International Conference on Multimedia and Expo., pp. 1082–1085 (2005)

11. Yin, L., Jia, J., Morrissey, J.: Towards race-related face identification: Research on skin color transfer. In: Proc. of IEEE Conf. on AFGR, pp. 362–368 (2004)
12. Hu, Y., Zhang, Z., Xu, X., Fu, Y., Huang, T.S.: Building Large Scale 3D Face Database for Face Analysis. In: Sebe, N., Liu, Y., Zhuang, Y.-t., Huang, T.S. (eds.) MCAM 2007. LNCS, vol. 4577, pp. 343–350. Springer, Heidelberg (2007)
13. Davidenko, N.: Silhouetted face profiles: A new methodology for face perception research. Journal of Vision 7, 1–17 (2007)
14. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. on Pattern Analysis and Machine Intelligence 24, 509–522 (2002)