GENDER AND ETHNICITY IDENTIFICATION FROM SILHOUETTED FACE PROFILES

Usman Tariq, Yuxiao Hu, Thomas S. Huang

Department of Electrical and Computer Engineering,
Coordinated Science Laboratory, and
Beckman Institute for Advanced Science and Technology,
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ABSTRACT

This paper demonstrates, to our best knowledge, the first attempt on gender and ethnicity identification from silhouetted face profiles using a computer vision technique. The results achieved, after testing on 441 images, show that silhouetted face profiles have a lot of information, in particular, for ethnicity identification. Shape context based matching [1] was employed for classification. The test samples were multi-ethnic. Average accuracy for gender was 71.20% and for ethnicity 71.66%. However, the accuracy was significantly higher for some classes, such as 83.41% for females (in case of gender identification) and 80.37% for East and South East Asians (in case of ethnicity identification).

Index Terms— Face recognition, Pattern recognition, Image processing

1. INTRODUCTION

Gender and Ethnicity identification present yet another challenge in face processing. These have an increasing number of applications as Human-Computer Interaction (HCI) and visual surveillance technologies evolve. Gender identification can be useful in face recognition, as this shall reduce the problem of matching the face with half of the database (provided both the genders have equal probability of occurrence in the database). Ethnicity identification shall reduce this problem even further (provided the database is multi-ethnic). In HCI, for instance, the computer can adapt to a person's sex in terms of processing (e.g.) the person's voice or offering the person the options which may be more specific and useful to a particular gender, etc. These may also aid shopkeepers; for example, to know the demographic distribution of the customers over a period of time.

In the following sections, we'll present the novelty of this paper, description of the database, experimental setup and finally the results and discussion. Please note that the terms race and ethnicity, in this paper, refer to a group of people who share similar facial features, which perceptually distinguish them from other groups (ethnicities).

2. BACKGROUND

Humans are very accurate in deciding the gender of a face even when cues from makeup, hairstyle and facial hair are minimized [2]. The results from [3] show that both the color and shape are vital in deciding the sex and race from a face by humans. Bruce et al. in [2] showed that the 'average' male face differs from an 'average' female face in the 3-D representation of the face (obtained by laser scanning), by having a more protuberant nose/brow and more prominent chin/jaw. Shape was also found important in race (ethnicity) decisions as shown in [3].

Davidenko in [4] reveals the presence of cues for humans, even in silhouetted face profiles, for gender identification. In a more recent study, Davidenko et al. [5] claimed that face silhouettes also have information for ethnicity identification.

Most of the existing work in the computer vision community uses frontal face images for gender or ethnicity identification. Some examples of the existing literature on gender identification may be found in [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16] and [17]; while for ethnicity identification may be found in [8], [11], [17], [18] and [19]. But none of the papers ever experimented with silhouetted face profiles.

The aim of this paper is to exploit the information available in the silhouetted profile faces reported in [4] and [5]. This paper, to our best knowledge, demonstrates the **first attempt** for gender and ethnicity identification using silhouetted face profiles with any computer vision technique. In future, this may be fused with some other method.

3. DATABASE

The silhouetted face profiles in our database were generated from the 3D face models collected by Hu et al. in [20]. We restricted the age of subjects in the database, used in our experiments, within the range 18-30 yrs to avoid any age-related distortions on the silhouettes. Males, whose beards and moustaches distorted their facial contours, were also not included.

The database had 441 images. It was divided into following four ethnic categories based upon the demographic information given by the participants.

1. *Black (B).* These contained Africans or African Americans
2. *East and Southeast Asian (ESEA).* These contain Chinese, Japanese, Korean, Vietnamese, Filipino, Singaporean, etc.
3. *South Asian (SA).* These include Indian, Pakistani, Sri-Lankan, Bangladeshi, etc.
4. *White (W).* These include Caucasian, Middle Eastern.

The demographic distribution of the database is given in Table 1. Some example silhouettes and their (part of) extracted profile contours (later used for processing and referred to as 'profile edge') are shown in Figure 1.

Ethnic and Gender Distribution of Database							
B		**W**		**ESEA**		**SA**	
67		225		107		42	
M	F	M	F	M	F	M	F
32	35	118	107	55	52	25	17

Table 1. Demographic distribution of the database

Fig. 1. Silhouetted face profiles from the database (*column 1 and 3*) and extracted portion of respective profile contours (*column 2 and 4*)

4. SHAPE CONTEXT

We used shape context based matching in our experiments, as the problem at hand was essentially to match contours. In this method, the shape of an object is represented by a discrete set of points sampled from the internal and external contours on the object. For a point p_i on a shape having n sampled points, the coarse histogram h_i of the coordinates of the remaining n-1 points, relative to p_i, is called the shape context of p_i [1]. Mathematically,

$$h_i(k) \;=\; \#\{q \neq p_i : (q - p_i) \in bin(k)\} \tag{1}$$

The bins used, are uniform in log-polar space, making shape context more sensitive to nearby points to p_i. The cost of matching the point p_i on first shape to the point q_j on the second shape, denoted by C_{ij} is given by the chi-square test statistic [1],

$$C_{ij} \equiv C(p_i, q_j) = \frac{1}{2}\sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \tag{2}$$

where $h_i(k)$ and $h_j(k)$ denote the k-bin normalized histogram at p_i and q_j respectively.

Given the costs C_{ij} between all pairs of point p_i q_j on the two shapes, the total cost of matching is given by [1],

$$H(\pi) = \sum_i C(p_i, q_{\pi(i)})) \tag{3}$$

We want to minimize $H(\pi)$ subject to the constraint that matching is one-to-one, in other words, π is a permutation [1]. This is solved using Hungarian [21].

5. COORDINATE TRANSFORMATION

The correspondence between points can be used to estimate a plane transformation T given by $T: R^2 \rightarrow R^2$. This may be an affine transformation. But we used the thin plate spline model (TPS) [22], [23]. It is commonly used for flexible coordinate transformation [1]. In its regularized form, it includes the affine model as a special case [1].

Two separate TPS functions to model a coordinate transformation were used [1],

$$T(x, y) = (f_x(x, y), f_y(x, y)) \tag{4}$$

where the TPS interpolant $f(x,y)$ minimizes the bending energy I_f and is given by [22],

$$f(x, y) = a_1 + a_x x + a_y + \sum_{i=1}^{n} \omega_i U(\|(x_i, y_i) - (x, y)\|) \tag{5}$$

where (x_i,y_i) = the x-, y- coordinates of the point p_i, and $i=1,2,\ldots,n$; kernel function $U(r) = r^2 \log r^2$, $U(0) = 0$; ω_i are weights; a_1, a_x, a_y are constants. Also for $f(x,y)$ to have square integrable derivates, it is required that [1] [22],

$$\sum_{i=1}^{n} \omega_i = 0,\ also\ \sum_{i=1}^{n} \omega_i x_i = \sum_{i=1}^{n} \omega_i y_i = 0 \tag{6}$$

Bending energy is given by [22],

$$I_f \propto \omega^T K \omega = D_{be} \tag{7}$$

where $K_{ij}= U(\|(x_i,y_i)-(x_j,y_j)\|)$. The exact interpolation requirement may be relaxed using regularization [1].

6. METHODOLOGY

We calculated a "shape distance" of the test profile edge to the training profile edges by an approach similar to the one described by Belongie et al. in [1] using shape contexts. The methodology goes as follows:

- The following step is not required for shape context based matching [1] but was done to avoid computations of the correspondences of points that lie on, say, forehead of one profile and chin of the other etc., thus to reduce the computation time in 'Hungarian'. The profile edges were made vertical and facing to the right, then their heights were made equal (300 pixels).

- Sample both the test profile edge and training profile edge by 100 points. This can be done uniformly at random, however it was found advantageous to have a certain minimum distance between the sampled points in [1]. This makes the sampling somewhat uniform.
- Calculate the shape context [1] of each point on both the profiles edges
- Compute the correspondence between points using Hungarian [21].
- Use the correspondence between points to estimate a plane transformation *T*, given by $T: R^2 \rightarrow R^2$, by thin plate spline.
- Calculate D_{be}, a measure of the bending energy I_f, by equation (7).
- Calculate the shape context distance, D_{sc}, between the profile edges *P* and *Q*, as the symmetric sum of the shape context matching costs over best matching points [1], as in equation (8).

$$D_{sc}(P,Q) = \frac{1}{n}\sum_{p\in P}\arg\min_{q\in Q} C(p,T(q)) + \frac{1}{m}\sum_{q\in Q}\arg\min_{p\in P} C(p,T(q)) \tag{8}$$

- Initial estimates of correspondences may contain some errors [1]. This can be overcome by iterating the steps of recovering correspondences and estimating transformations. We experimented with 1, 2… 6 iterations.
- Calculate the *shape distance*, D_{sh}, as a weighted sum of D_{sc} and D_{be}.

$$D_{sh} = w_{sc} \times D_{sc} + w_{be} \times D_{be} \tag{9}$$

- Calculate shape distances, D_{sh}, of a test profile with all the training profile edges, and do classification using *k-Nearest Neighbors*.

7. EXPERIMENTS AND RESULTS

We made the training database equally balanced in terms of each gender-ethnic category (male Black, female Black, male South Asian and so on…), as we did not want the prior probabilities to affect the result of the Nearest Neighbor classification. The training dataset had 128 images.

First the remaining dataset was tested on training database. Then we used a 'replace-one approach' to test the members of training dataset. For this, each member in the training database we replaced, one-by-one, with an image of the same ethnicity and gender and then the algorithm was executed. In this way, the approach was tested on a total of 441 images.

The amounts of weights in equation (9) and iterations were empirically determined. The resulting accuracies achieved are summarized in Table 2. The class confusion matrices for the results in Table 2, is given for gender in Table 3 and for ethnicity in Table 4 and Table 5. For explanation of the abbreviations used for ethnicities, please refer to section 3. We also included the notion of Standard Error ($SE = \sqrt{p(1-p)/n}$, where p is the accuracy) in our computations.

	Gender	Standard Error	**Ethnicity**	Standard Error
Accuracy	71.20%	± 2.16%	71.66%	± 2.15%

Table 2. Results achieved

Ground Truth	**Recognized As**		**Recognized As**			
			Female		Male	
	Female	Male	%	SE %	%	SE %
Female	176	35	**83.41**	± 2.56	16.59	± 2.56
Male	92	138	40	± 3.23	**60**	± 3.23

Table 3. Confusion matrix for gender

Ground Truth	**Recognized As**			
	B	ESEA	SA	W
B	39	24	0	4
ESEA	7	86	6	8
SA	0	13	14	15
W	1	12	35	177

Table 4. Confusion matrix for Ethnicity

Ground Truth	**Recognized As**							
	B		ESEA		SA		W	
	%	SE %	%	SE %	%	SE %	%	SE %
B	**58.21**	± 6.0	35.82	± 5.9	0.00	0.0	5.97	± 2.9
ESEA	6.54	± 2.3	**80.37**	± 3.8	5.61	± 2.2	7.48	± 2.5
SA	0.00	0.0	30.95	± 7.1	33.33	± 7.3	35.71	± 7.4
W	0.44	± 0.4	5.33	± 1.5	15.56	± 2.4	**78.67**	± 2.7

Table 5. Confusion matrix for Ethnicity (percentages)

8. OBSERVATIONS AND DISCUSSION

It may be observed that females seem to be more reliably identified than males, evident by their accuracy of (83.41 ± 2.56)%, shown in Table 3. This is in contrast to the frontal views, in which generally the bias is towards males in performance, as also noted in [11].

Eeast and South East Asians seems to be the most reliably identified ethnic category, evident by their accuracy of (80.37 ± 3.8)%, shown in Table 5. Whites and Blacks also share a higher accuracy of true identification. Blacks are generally confused with East and South East Asians. South Asians are generally confused both with Whites and East and South East Asians. Whites are occasionally confused with South Asians.

The results were also found superior to the average human performance for silhuoetted face profiles, noticed in an experiment, done by our group, on 21 subjects. The average performance in that experiment was 57.63% (std. dev.=9.43%) for gender and 45.08% (std. dev.=18.75%) for ethnicity identification. The effect of resolution on performance, however, had not been experimented with in this work. This method should work fine if the resolution is reasonable enough to capture the variations in silhouettes.

We have demosntrated the first attempt for application of a computer vision technique in gender and ethnicity identification from silhuoetted face profiles. This may be fused with some other face view in future. This, most probably, would improve performance as silhuoetted face profiles have a lot of information as shown in this paper.

9. ACKNOWLEDGMENTS

This research is supported in part by U. S. Government VACE Program.

This paper has had the benefit of some very useful suggestions from Prof. David Forsyth.

In this project, the Matlab implementation of Hungarian by Niclas Borlin was used.

10. REFERENCES

[1] S. Belongie, J. Malik, J. Puzicha, "Shape matching and object recognition using shape contexts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 509-522, 2002.

[2] V. Bruce, A. M. Burton, E. Hanna, P. Healey, O. Mason, A. Coombes, R. Fright, A. Linney, "Sex discrimination: how do we tell the difference between male and female faces?", *Perception*, vol. 22, pp. 131-152, 1993.

[3] H. Hill, V. Bruce, S. Akamatsu, "Perceiving the sex and race of faces: The role of shape and color". *Proceedings of the Royal Society - Biological Sciences (Series B)*, vol. 261, pp. 367-373, 1995.

[4] N. Davidenko, "Silhouetted face profiles: A new methodology for face perception research", *Journal of Vision*, vol. 7, pp. 1-17, 2007.

[5] N. Davidenko, N. Witthoft, J. Winawer, "Gender aftereffects in face silhouettes reveal face-specific mechanisms", *Visual Cognition*, vol. 16, pp. 99-103, 2008.

[6] B.A. Golomb, D.T. Lawrence, T.J. Sejnowski, "Sexnet: a neural network identifies sex from human faces", in *Advances in Neural Information Processing Systems (NIPS)*, vol. 3, pp. 572-577, Morgan Kaufmann, San Mateo, CA, 1990.

[7] R. Brunelli, T. Poggio, "Hyperbf networks for gender classification", *Proceedings of the Image Understanding Workshop*, pp. 311-314, 1992.

[8] S. Gutta, J.R.J. Huang, P. Jonathon, H. Wechsler, "Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces", *IEEE Transactions on Neural Networks*, vol. 11, pp. 948-960, 2000.

[9] B. Moghaddam, M. Yang, "Learning gender with support faces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 707-711, 2002.

[10] A.B.A. Graf, F.A. Wichmann, "Gender classification of human faces", in *Biologically Motivated Computer Vision 2002, Proceesdings*. LNCS, vol. 2525, pp. 491-500, Springer, Heidelberg, 2002.

[11] G. Shakhnarovich, P.A. Viola, B. Moghaddam, "A unified learning framework for real time face detection and classification", in *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp. 14-21, 2002.

[12] Z. Sun, G. Bebis, X. Yuan, S.J. Louis, "Genetic feature subset selection for gender classification: a comparison study", in *IEEE Workshop on Application of Computer Vision*, pp. 165-170, 2002.

[13] L. Walawalkar, M. Yeasin, A. Narasimhamurthy, R. Sharma, "Support vector learning for gender classification using audio and visual cues", *International Journal on Pattern Recognition and Artificial Intelligence*, vol. 17, pp. 417-439, 2003.

[14] A. Jain, J. Huang, "Integrating independent component analysis and linear discriminant analysis for gender classification", in *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp. 159-163, 2004.

[15] X. Xu, T.S. Huang, "SODA-Boosting and Its Application to Gender Recognition", in *Analysis and Modeling of Faces and Gestures 2007, Proceedings*. LNCS, vol. 4778, pp. 193-204, Springer Heidelberg, 2007.

[16] C.F. Shan, S.G. Gong, P.W. McOwan, "Fusing gait and face cues for human gender recognition", *Neurocomputing*, vol. 71, pp. 1931-1938, 2008.

[17] X.G. Lu, H. Chen, A.K. Jain, "Multimodal Facial Gender and Ethnicity Identification", in *Advances in Biometrics*. LCNS, vol. 3832, pp. 554-561, Springer, Heidelberg, 2005.

[18] A. J. O'Toole, H. Abidi, K.A. Deffenbacher, J.C. Bertlett, "Classifying faces by race and sex using an autoassociative memory trained for recognition", in *Hammond, K. J., Genter, D. (eds.) Proc. 13th Annu. Conf. Cognitive Sci. Soc.*, Lawrence Erlbaum, NJ, 1991.

[19] X. Lu, A.K. Jain, "Ethnicity identification from face images", in *Biometric Technology for Human Identification. Proceedings of SPIE - The International Society for Optical Engineering*, vol. 540, pp. 114-123, 2004.

[20] Y. Hu, Z. Zhang, X. Xu, Y. Fu, T.S. Huang, "Building Large Scale 3D Face Database for Face Analysis", in *Multimedia Content Analysis and Mining* 2007. LNCS, vol. 4577, pp. 343-350, Springer, Heidelberg, 2007.

[21] C. Papadimitriou, K. Stieglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice Hall, 1982.

[22] F.L. Bookstein, "Principal Warps - Thin-Plate Splines and the Decomposition of Deformations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 567-585, 1989.

[23] J. Meinguet. "Multivariate Interpolation at Arbitrary Points Made Simple", *Journal of Applied Math and Physics (ZAMP)*, vol. 30, pp. 292-304, 1979.

[24] E. Brown, D. I. Perrett, "What gives a face its gender?", *Perception*, vol. 22, pp. 829 - 840, 1993.

[25] L. Wiskott, "Phantom faces for face analysis", *Pattern Recognition*, vol. 30, pp. 837-847, 1997.

[26] V. Blanz, T. Vetter. "A Morphable Model for the Synthesis of 3D Faces", in *Proc. Siggraph99*, pp. 187-194, ACM Press, Los Angeles, 1999.

[27] F.A. Wichmann, A.B.A. Graf, E.P. Simoncelli, H.H. Bulthoff, B. Scholkopf, "Machine learning applied to perception: decision-images for gender classi-fication", in *Advances in Neural Information Processing Systems (NIPS)*, 2004.