Investigating the Impact of Face Categorization on Recognition Performance

Konstantinos Veropoulos[1], George Bebis[1], and Michael Webster[2]

[1]Computer Vision Laboratory
[2]Department of Psychology
University of Nevada, Reno NV 89557, USA
{kvero,bebis}@cse.unr.edu, mwebster@unr.edu

Abstract. Face recognition is a key biometric technology with a wide range of potential applications both in government and private sectors. Despite considerable progress in face recognition research over the past decade, today's face recognition systems are not accurate or robust enough to be fully deployed in high security environments. In this paper, we investigate the impact of face categorization on recognition performance. In general, face categorization can be used as a filtering step to limit the search space during identification (e.g., a person categorized as a middle-aged, Asian male, needs to be compared only to subjects having the same profile). Our experimental results demonstrate that face categorization based on important visual characteristics such as gender, ethnicity, and age offers significant improvements in recognition performance including higher recognition accuracy, lower time requirements, and graceful degradation. Additional performance improvements can be expected by implementing "category-specific" recognition subsystems that are optimized to discriminate more accurately between faces within the same face category rather than faces between other categories.

1 Introduction

Recently, there has been an increased interest in developing computer vision systems that can robustly and reliably recognize, track, monitor, and identify people and interpret their actions. Face recognition is a key biometric technology with a wide range of potential applications. Despite considerable progress in this research area, today's face recognition systems are not accurate or robust enough to be fully deployed in high security environments. Advances in this area are thus likely to make significant contributions in areas such as security, monitoring, surveillance and safety. Motivated by cognitive evidence, we believe that significant gains in recognition performance can be achieved by applying face categorization prior to recognition and optimizing recognition within each face category.

Specifically, there is cognitive evidence supporting the idea that humans utilize information from multiple visual cues for face recognition. It is well known, for example, that people are more accurate at recognizing faces of their own ethnicity than faces of another ethnicity [1]. Humans can also judge the gender of adults and children using feature sets derived from the appropriate face age category, rather than applying features derived from another age category or a combination of age categories [2]. It has been also found that adaptation may routinely influence face perception and could have an important role in calibrating properties of face perception according to the subset of faces populating an individual's environment [3]. All this evidence suggests that people are better skilled than machines in recognizing faces because they have developed "specialized" perceptual processes through lifelong experiences while adaptation self-calibrates the human vision system to faces in their environment. These processes and the ability of the human visual system to adapt allow them to be more sensitive to certain types of visual information (e.g., age or gender), carrying more discriminative power for faces within the same category.

Despite the significant amount of evidence in this area, typical face recognition systems do not exploit, at least explicitly, information from multiple visual cues for recognition. In fact, in most cases faces are represented by extracting the same type features regardless to differences in gender, ethnicity, and age (e.g., middle-aged, Asian male vs young, Black female). Therefore, it is reasonable to expect that face recognition suffers from irrelevant and/or redundant information. In addition, many times face recognition systems yield inconsistent matches (e.g., matching a middle-aged, Asian male to a young, Black female). In principle, inconsistent matches can be avoided or reduced by simply restricting search only to faces having the same profile with the face in question.

Although some attention has been given to the problem of gender [4][5], ethnicity [6] and age classification [7], typical face recognition systems do not explicitly exploit information from such visual cues to limit the search space and reduce the number of inconsistent matches. An exemption is the recent work of Jain *et al.* [8] which shows that using ancillary information based on "soft biometrics" (e.g., gender and ethnicity) leads to improving the recognition performance of a fingerprint system. However, they have used this information along with the output of the fingerprint recognition system in order to verify the matching results, rather than exploiting this information to reduce the search space prior matching.

Our emphasis is this work is on investigating the impact of face categorization on recognition performance. In this context, we have designed and performed a large number of experiments using the FERET database to demonstrate the benefits of applying face categorization prior to recognition. Additional performance improvements can be expected by designing "specialized" (i.e., category-specific) recognition subsystems that are explicitly optimized to discriminate more accurately between faces in the same face category than faces in other categories. Coupling face categorization with category-specific recognition is essentially equivalent to incorporating an adaptation mechanism to the recognition process, allowing recognition to self-calibrate itself to different types of faces in the operating environment. Although we are not dealing here with the design and implementation of face categorization and category-specific recognition, we do discuss in Section 6 a number of important issues.

The rest of the paper is organized as follows: Section 2 presents a general methodology for coupling face categorization with recognition. Section 3 briefly describes the face recognition approach used in this study. The datasets and our evaluation methodology are presented in Section 4 while experimental results are presented in Section 5. Finally, in Section 6, we analyze the results of our experiments and elaborate on the design and implementation of face categorization and category-specific recognition algorithms.

2 Methodology

The key idea of employing face categorization is dividing faces into different categories prior to recognition using information from various visual cues. *First,* the face database (i.e., gallery set) is divided into different subsets by assigning faces into different face categories using gender, ethnicity and/or age information. It should be noted that, the purpose of age classification is to assign a given face to a particular age group (eg. between 20 to 40 years old) rather than estimating the age of the subject exactly. *Second*, a given face (i.e., query) is assigned to the appropriate face category using the same procedure. The query is then matched against faces belonging to its assigned category only, instead of being compared to all the faces in the database. A simple diagram illustrating the above procedure is shown in Fig. 1.

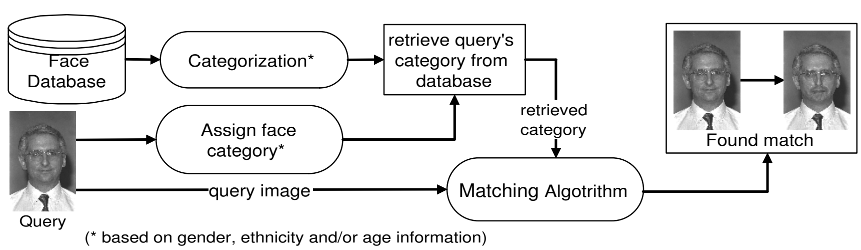


Fig. 1. Categorization and matching procedure for a query image.

Applying categorization prior to recognition restricts the search space from the whole face database to the subset of images that belong to the same face category as the query image. This way, not only the number of comparisons is reduced to the size of the chosen face category, thus speeding up the matching process, but also the risk of mismatching a given face to a face from a completely different face category is reduced. In other words, it allows the system to degrade gracefully.

3 Recognition

To quantify the effects of face categorization on recognition, we performed recognition using the popular method of eigenfaces [9], although any other recognition methodology could have been used. The eigenface approach uses Principal Component Analysis (PCA) to represent faces in a low-dimensional subspace spanned by the "largest" eigenvectors of the covariance matrix of the data, that is, the eigenvectors corresponding to the largest eigenvalues or the directions of maximum variance. To create the eigenspace, 400 images were randomly chosen from the gallery set (see next section). In choosing the largest eigenvectors, we preserve 97% of the information in the data. For matching, we used a minimum-distance classifier based on Mahalanobis distance. Given a query face, we find the top N faces having the highest similarity score with the query face. To evaluate matching, we used the Cumulative Match Characteristic (CMC) curve [10] which shows the probability of identification against the returned 1-to-N candidate list size (i.e., it shows the probability that a given face appears in different sized candidate lists). The faster the CMC curve approaches the value one, which is an indication of that face being in the candidate list of a specified size, the better the matching algorithm is.

4 Datasets and Evaluation

To test our approach, we used the FERET database [10], released in October 2003, which contains a large number of images acquired during different photo sessions and has a good variety of gender, ethnicity and age groups. The lighting conditions, face orientation and time of capture vary. In this work, we concentrate on frontal face poses coded as *fa* (regular frontal image) or *fb* (alternative frontal image, taken *shortly* after the corresponding *fa* image)[1]. In our evaluations, the *fa* images were used as the gallery set while the *fb* images were used as the query set (i.e., face images in question). All faces were normalized in terms of orientation, position and size prior to experimentation. They were also masked to include only the face region (i.e., upper body and background were cropped out) yielding an image size of 48×60 pixels.

In order to evaluate the effect of different levels of face categorization, three sets of experiments were designed. First, we wanted to see what kind of improvements could be expected using one-level categorization where only one type of information was used to categorize faces (i.e., gender, ethnicity, or age). Next, we investigated a two-level categorization where two types of information were used at a time (i.e., gender and ethnicity, gender and age, or ethnicity and age). Finally, we investigated a three-level categorization where all three types of information (i.e., gender, ethnicity, and age) were used.

Before proceeding in describing our experiments and presenting our results, two important issues must be clarified. First, in practice, face categorization

[1] It should be noted that, *shortly* could mean up to two years after the *fa* pose had been taken.

will be an automated process. In this study, however, face categorization was implemented manually since our main objective was to investigate the impact of categorization on recognition performance. Therefore, the results presented here can be thought as "best-case" performance since we have assumed error-free categorization. Second, recognition has not been optimized for each face category (i.e., we have applied the same recognition procedure for each face category). In general, one can expect higher performance by optimizing recognition within each face category. There are important issues to be considered in both cases which we discuss in Section 6. In the following subsections, we provide a detailed description of our experiments.

One-Level Categorization: In one-level categorization, we only consider one category at a time, that is, gender, ethnicity or age separately. This type of categorization results in three different groups of experiments with the data organization shown in Table 1. Our notation for ethnicity categorization is as follows: Asian (As), Asian-Middle-Eastern (AME), Black-or-African-American (BAA), Hispanic (Hisp) and White (Wh).

Table 1. Organization of data (total number of persons and, in parentheses, total number of images) for the gallery set (fa) and the query set (fb) assuming one-level partitioning.

	Gender		**Ethnicity**				
	Male	**Female**	**As**	**AME**	**BAA**	**Hisp**	**Wh**
fa	501 (746)	365 (457)	130 (192)	40 (61)	72 (100)	51 (63)	558 (770)
fb	500 (740)	366 (456)	130 (190)	40 (60)	72 (99)	51 (63)	558 (767)

	Age					
	07–13	**17–23**	**27–33**	**37–43**	**46–53**	**57–63**
fa	17 (18)	402 (484)	201 (296)	144 (251)	79 (120)	21 (30)
fb	17 (18)	403 (480)	200 (296)	144 (247)	79 (120)	21 (31)

Two-Level Categorization: In two-level categorization, we consider combinations of gender and ethnicity, gender and age, and ethnicity and age. This type of categorization results in three different groups of experiments with the data organization shown in Table 2. It should be noted that all categories including a small number of subjects (i.e., less than 10) have not been included in our evaluations (e.g., Hispanics between 27 and 33 years of age).

There-Level Categorization: In three-level categorization, we consider all three types of information (gender, ethnicity and age) together. This type of categorization results in several groups of experiments with the data organization shown in Table 3. Again, all groups containing less than 10 subjects have not been included in our evaluations.

Table 2. Organization of data (total number of persons and, in parentheses, total number of images) for the gallery set (fa) and the query set (fb) after a two-level partitioning.

	Gender+Ethnicity			
	fa		**fb**	
	Male	**Female**	**Male**	**Female**
As	83 (134)	47 (58)	82 (131)	48 (59)
AME	35 (53)		35 (53)	
BAA	33 (47)	39 (53)	33 (46)	39 (53)
Hisp	23 (28)	28 (35)	23 (28)	28 (35)
Wh	321 (478)	237 (292)	321 (476)	237 (291)

	Gender+Age			
	fa		**fb**	
	Male	**Female**	**Male**	**Female**
07–13	10 (11)		10 (11)	
17–23	183 (239)	219 (245)	183 (235)	220 (245)
27–33	131 (239)	70 (97)	130 (199)	70 (97)
37–43	102 (181)	42 (70)	102 (178)	42 (69)
47–53	55 (86)	24 (34)	55 (86)	24 (34)
57–63	19 (28)		19 (29)	

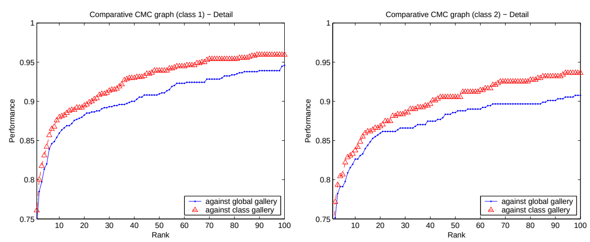
	Ethnicity+Age									
	fa					**fb**				
	As	**AME**	**BAA**	**Hisp**	**Wh**	**As**	**AME**	**BAA**	**Hisp**	**Wh**
07–13					10 (11)					10 (11)
17–23	72 (93)	19 (34)	31 (36)	31 (34)	241 (277)	73 (92)	19 (33)	31 (36)	31 (34)	241 (275)
27–33	41 (69)	10 (11)	15 (22)		123 (177)	40 (68)	10 (11)	15 (22)		123 (178)
37–43			17 (30)		104 (179)			17 (30)		104 (176)
47–53					58 (93)					58 (93)
57–63					29 (20)					20 (30)

Table 3. Organization of data (total number of persons and, in parentheses, total number of images) for the gallery set (fa) and the query set (fb) after a three-level partitioning.

	Gender+Ethnicity+Age			
	fa		**fb**	
	Male	**Female**	**Male**	**Female**
As/17-13	33 (47)	39 (46)	33 (45)	40 (47)
As/27-33	35 (61)		34 (60)	
AME/17-13	14 (26)		14 (26)	
AME/27-33	10 (11)		10 (11)	
BAA/17-13	11 (13)	20 (23)	11 (13)	20 (23)
BAA/37-43	10 (16)		10 (16)	
Hisp/17-13	15 (17)	16 (17)	15 (17)	16 (17)
Wh/17-13	110 (136)	131 (141)	110 (134)	131 (141)
Wh/27-33	73 (109)	50 (68)	73 (110)	50 (68)
Wh/37-43	72 (128)	32 (51)	72 (126)	32 (50)
Wh/47-53	40 (68)	18 (25)	40 (68)	18 (25)
Wh/57-63	18 (27)		18 (28)	

5 Experimental Results

To quantify the effect of face categorization, we compared the CMC curves in two cases: (a) when search is restricted to a particular face category (best curve) and (b) when searching the whole gallery set (worst curve). As mentioned earlier, when the CMC curve reaches the value one, then the face in question (query) always appears in the candidate list. For example, the CMC curve for Asian-Middle-Eastern faces (class 2) shown in Fig. 3, indicates that the query face always appears in a candidate list of at least 27 persons assuming categorization, whereas to have the same effect without categorization, the candidate list needs to contain more than 100 persons.



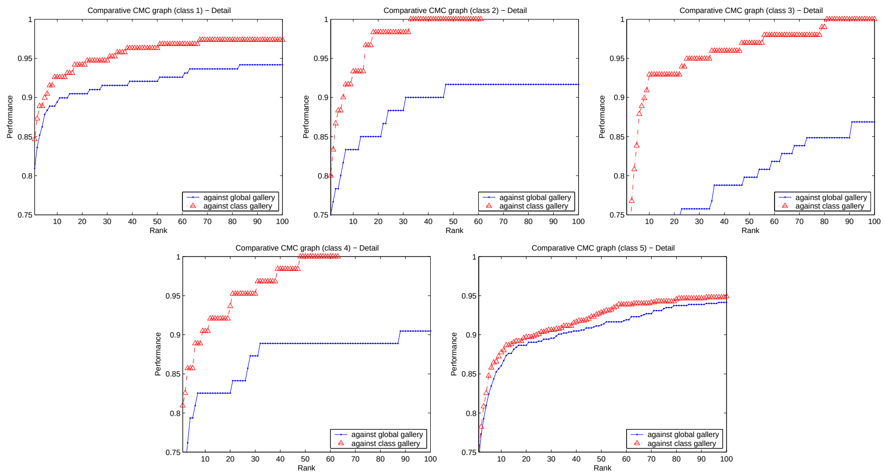
Key:	Class 1: Male
	Class 2: Female

Fig. 2. CMC curves for Males and Females (one-level categorization).

Figs. 2–4 present the CMC plots obtained from experiments on one-level categorization. CMC plots obtained from experiments on two-level categorization are shown in Figs. 6–8. Finally, the CMC plots obtained from three-level categorization are given in Fig. 9. Fig. 5 shows several test cases where recognition failed when comparing each of the test faces against the whole gallery set (i.e., without face categorization). The purpose of the example is to demonstrate what kind of matching errors one should expect (e.g., mismatching a male/BBA (#854) to a female/Hispanic (#351)). Intuitively, one would expect recognition to degrade gracefully, that is, mismatching people within the same face category but not between different categories. Coupling face categorization with recognition has the potential to reduce the number of inconsistent matches.

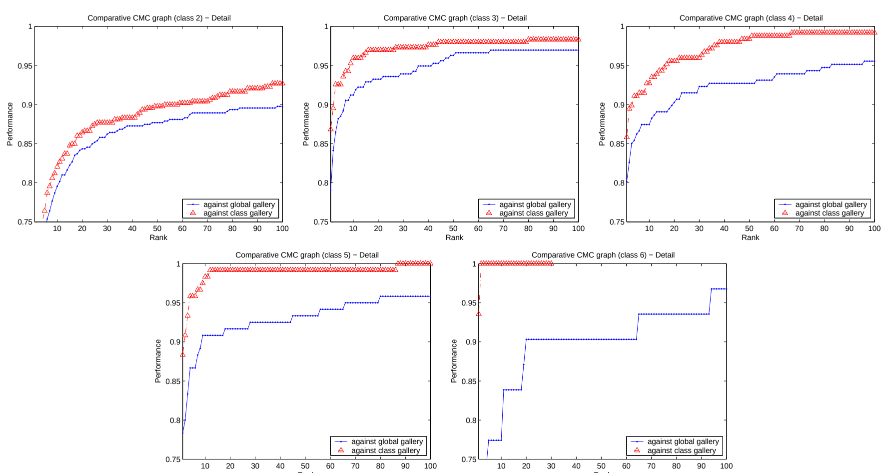
6 Discussion and Conclusions

All CMC plots in Figs. 2–9 illustrate that applying face categorization prior to recognition leads to recognition improvements by reducing the search space and increasing accuracy. It is worth mentioning that, face categorization is independent of the recognition algorithm, therefore, it could be coupled with existing



Key:			
	Class 1: As	Class 3: BAA	Class 5: White
	Class 2: AME	Class 4: Hisp	

Fig. 3. CMC curves for different ethnicities (one-level categorization).



Key:			
	Class 1: 07–13 years old	Class 3: 27–33 years old	Class 5: 47–53 years old
	Class 2: 17–23 years old	Class 4: 37–43 years old	Class 6: 57–63 years old

Fig. 4. CMC curves for different age groups (one-level categorization).

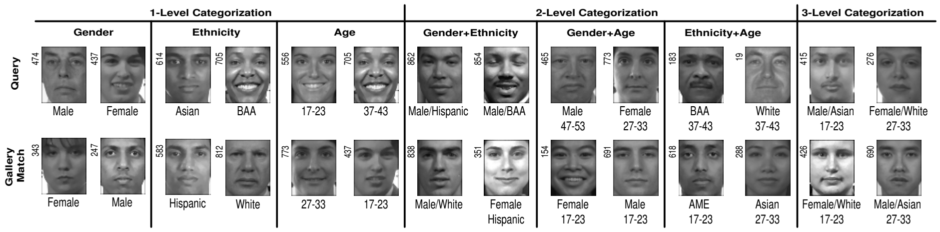
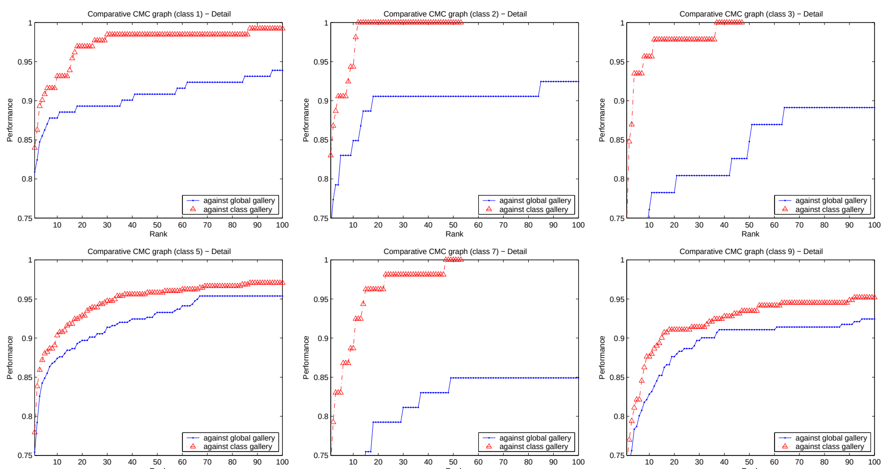


Fig. 5. Examples of mismatches without assuming face categorization. These cases were correctly matched assuming category-specific galleries.



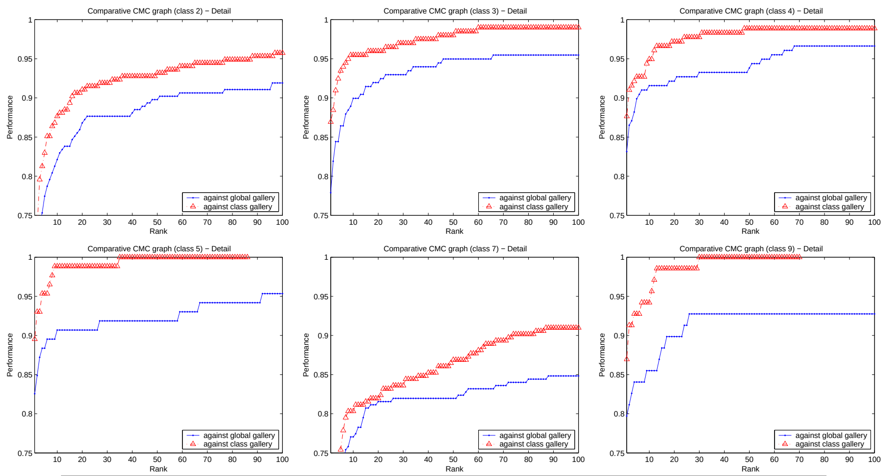
Key:			
	Class 1: Male/As	Class 4: Male/Hisp	Class 7: Female/BAA
	Class 2: Male/AME	Class 5: Male/Wh	Class 8: Female/Hisp
	Class 3: Male/BAA	Class 6: Female/As	Class 9: Female/Wh

Fig. 6. CMC curves for two-level categorization using gender and ethnicity information.

recognition systems without requiring any changes to the recognition engine or radical and costly changes in the current infrastructure.

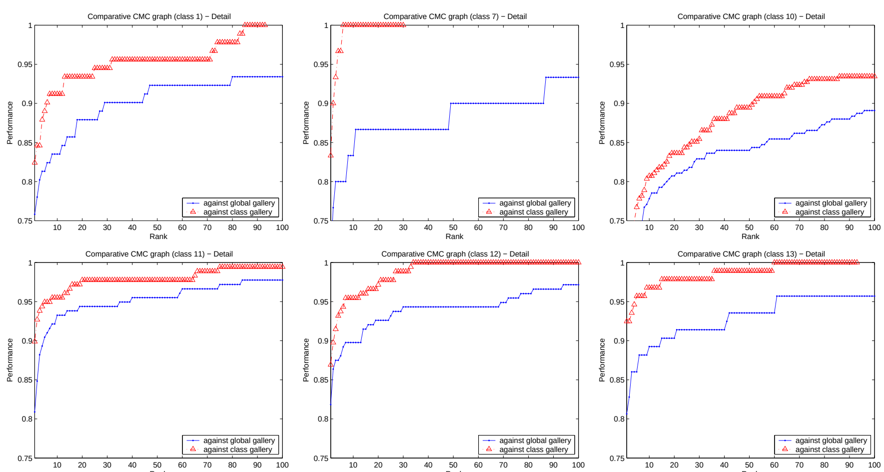
There are a number issues related to implementing face categorization. One of them is what cues to select to define the face categories. Another one is whether to perform "hard" or "soft" categorization. "Hard" categorization implies assigning a face to a single face category while "soft" categorization implies assigning a face to several categories, each with a certain probability. In this study, we used some of the most obvious visual cues (e.g., gender, ethnicity, and age), and "hard" categorization. However, it might be possible to use additional visual cues (e.g., face shape) or even cues that do not necessarily have an obvious visual interpretation (e.g., generate the face categories using unsupervised learning [11]).

An other issue is the error introduced by the categorization step. As discussed earlier, the results presented in this study assume error-free face categorization



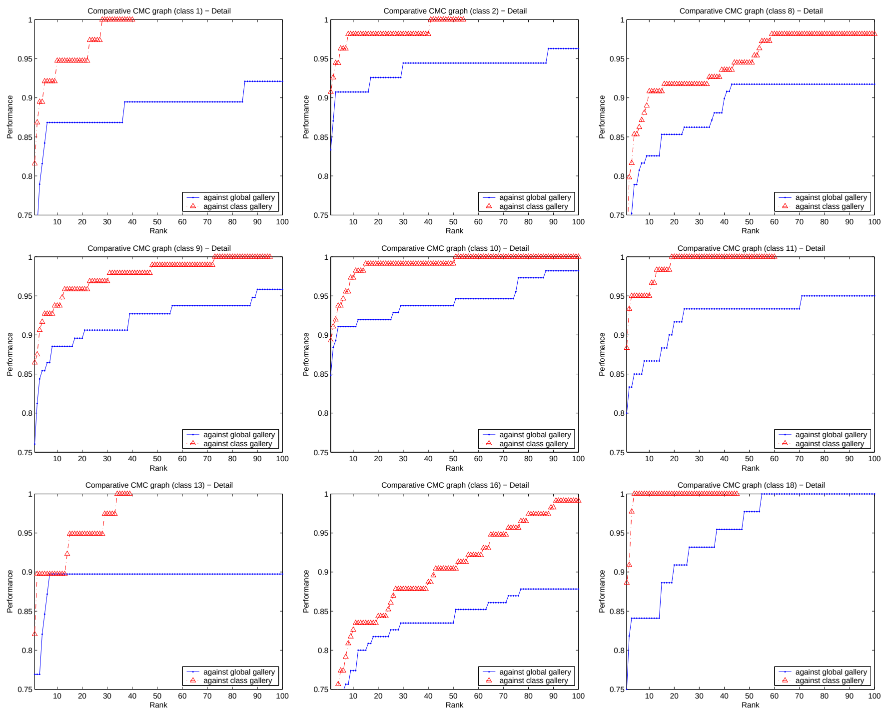
Key:				
	Class 1: Male/07–13	Class 4: Male/37–43	Class 7: Female/17–23	Class 10: Female/47–53
	Class 2: Male/17–23	Class 5: Male/47–53	Class 8: Female/27–33	
	Class 3: Male/27–33	Class 6: Male/57–63	Class 9: Female/37–43	

Fig. 7. CMC curves for two-level categorization using gender and age information.



Key:					
	Class 1: As/17-23	Class 4: AME/27-33	Class 7: BAA/37-43	Class 10: Wh/17-23	Class 13: Wh/47-53
	Class 2: As/27-33	Class 5: BAA/17-23	Class 8: Hisp/17-23	Class 11: Wh/27-33	Class 14: Wh/57-63
	Class 3: AME/17-23	Class 6: BAA/27-33	Class 9: Wh/07-13	Class 12: Wh/37-43	

Fig. 8. CMC curves for two-level categorization using ethnicity and age information.



Key:			
Class 1: Male/As/17–23	Class 6: Male/BAA/37–43	Class 11: Male/Wh/47–53	Class 16: Female/Wh/17–23
Class 2: Male/As/27–33	Class 7: Male/Hisp/17–23	Class 12: Male/Wh/57–63	Class 17: Female/Wh/27–33
Class 3: Male/AME/17–23	Class 8: Male/Wh/17–23	Class 13: Female/As/17–23	Class 18: Female/Wh/37–43
Class 4: Male/AME/27–33	Class 9: Male/Wh/27–33	Class 14: Female/BAA/17–23	Class 19: Female/Wh/47–53
Class 5: Male/BAA/17–23	Class 10: Male/Wh/37–43	Class 15: Female/Hisp/17–23	

Fig. 9. CMC curves for three-level categorization using gender, ethnicity and age information.

(i.e., performed manually). In practice, however, face categorization is expected to introduce some errors by assigning faces to wrong categories, leading to incorrect matches. Employing "soft" categorization instead of "hard" categorization could help to reduce these errors. In general, however, it would be necessary to design highly accurate and robust face categorization algorithms in order to achieve high recognition accuracy. We believe that one way to deal with this issue is by capitalizing on recent advances in pattern recognition and machine learning.

First, we believe that it would be important to optimize the face representation scheme used for each category. Let us take, for example, the case of PCA that was used here to represent faces. For each face category, we chose a subset of eigenvectors by applying the same principle (i.e., choosing the "largest" eigenvectors). Although the "largest" eigenvectors preserve most of the information in the data, it is well known that they might not provide the best possible discrimination power. Therefore, it would be essential to optimize the face representation scheme for each category by selecting "category-specific" eigenvectors. This is essentially equivalent to performing feature selection. We have done preliminary work on eigenvector selection for gender classification [5], showing that it is possible to improve gender classification by selecting eigenvectors that encode mostly gender information. Alternatively, it might be more appropriate to consider other representation schemes or combinations of them such as Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Canonical Correlation Analysis (CCA) [11] and their kernel counterparts [12].

Second, it would be important to employ more powerful classification algorithms such as Support Vector Machines (SVMs) and kernel methods [12]. Similar arguments can be made for the design of category-specific recognition subsystems. In addition to the above issues, there are also other issues such as how to deal with face categories containing a small number of subjects. As mentioned in Section 4, certain face categories in the FERET database contain less than 10 subjects. Training a classifier on a very small dataset becomes problematic and requires careful consideration. Our future work involves dealing with these issues.

References

1. O'Toole, A., Peterson, J., Deffenbacher, K.: An other-race effect for classifying faces by sex. In: Perception. Volume 25. (1996) 669–676
2. Cheng, Y., O'Toole, A., Abdi, H.: Classifying adults' and children's faces by sex: Computational investigations of subcategorical feature encoding. In: Cognitive Science. Volume 25. (2001) 819–838
3. Webster, M., Kaping, D., Mizokami, Y., Duhamel, P.: Adaptation to natural facial categories. In: Perception. Volume 428. (2004) 558–561
4. Moghaddam, B., Yang, M.: Learning gender with support faces. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 24. (2002) 707–711
5. Sun, Z., Bebis, G., Yuan, X., Louis, S.J.: Genetic feature subset selection for gender classification: A comparison study. In: IEEE Workshop on Applications of Computer Vision, Orlando, FL (2002) 165–170
6. Gutta, S., Huang, J., Phillips, J., Wechsler, H.: Mixture of experts for classification of gender, ethnic origin, and pose of human faces. In: IEEE Transactions on Neural Networks. Volume 11. (2000) 948–960
7. Kwon, Y.H., Lobo, N.d.V.: Age classification from facial images. In: Computer Vision and Image Understanding. Volume 74. (1999) 1–21
8. Jain, A.K., Dass, S.C., Nandakumar, K.: Soft biometric traits for personal recognition systems. In: Lecture Notes in Computer Science. Volume 3072. (2004) 731–738
9. Turk, M., Pentland, R.: Eigenfaces for recognition. Cognitive Neuroscience **3** (1991) 71–86
10. Phillips, J., Moon, H., Risvi, S., Rauss, J.: The feret evaluation methodology for face recognition algorithms. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 22. (2000) 1090–1104
11. R. Duda, P.H., Stork, D.: Pattern Classification. Jon-Wiley, 2nd edition (2001)
12. Taylor, J., Cristianini., N.: Kernel Methods for Pattern Analysis. Cambridge University Press (2004)