Short Papers

Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics

Kyong Chang, Kevin W. Bowyer, *Fellow*, *IEEE*, Sudeep Sarkar, *Member*, *IEEE*, and Barnabas Victor

Abstract—Researchers have suggested that the ear may have advantages over the face for biometric recognition. Our previous experiments with ear and face recognition, using the standard principal component analysis approach, showed lower recognition performance using ear images. We report results of similar experiments on larger data sets that are more rigorously controlled for relative quality of face and ear images. We find that recognition performance is not significantly different between the face and the ear, for example, 70.5 percent versus 71.6 percent, respectively, in one experiment. We also find that multimodal recognition using both the ear and face results in statistically significant improvement over either individual biometric, for example, 90.9 percent in the analogous experiment.

Index Terms—Biometrics, multimodal biometrics, face recognition, ear recognition, appearance-based recognition, principal component analysis.

1 Introduction

While good face recognition performance has been reported under certain conditions, there is still a great need for better performance in biometrics appropriate for use in video surveillance. Possible avenues for improved performance include the use of a different source of biometric information, and/or the combination of information from multiple sources. One other possible biometric source is the ear. Iannarelli performed important early research on a manual approach to using the ear for human identification [1]. Recent works that explore computer vision techniques for ear biometrics include those of Burge and Burger [2] and Hurley et al. [3]. In particular, Burge and Burger assert that the ear offers the promise of similar performance to the face:

> Facial biometrics fail due to the changes in features caused by expressions, cosmetics, hair styles, and the growth of facial hair as well as the difficulty of reliably extracting them in an unconstrained environment exhibiting imaging problems such as lighting and shadowing....Therefore, we propose a new class of biometrics for passive identification based upon ears which have both reliable and robust features which are extractable from a distance...identification by ear biometrics is promising because it is passive like face recognition, but instead of the difficult to extract face biometrics, robust and simply extracted biometrics like those in fingerprints can be used. ([2], p. 275)

In the context of Iannarelli's earlier work and the current popularity of face recognition research, this assertion that the ear could offer improved biometric performance relative to the face deserves careful evaluation. The experiments reported in this paper are aimed at 1) testing the hypothesis that *images* of the ear provide better biometric performance than *images* of the face and 2) exploring whether a combination of ear and face images may provide better performance than either one individually. The results reported here follow up on those reported in an earlier study [4]. Using larger data sets and more rigorous assurance of similar relative quality in the ear and face images, we obtain somewhat different results than in the earlier study. In the experiments reported here, recognition performance is essentially identical using ear images or face images and combining the two for multimodal recognition results in a statistically significant performance improvement. For example, in one experiment the rank-one recognition rates for face and ear were 70.5 percent and 71.6 percent, respectively, whereas the corresponding multimodal recognition rate was 90.9 percent. To our knowledge, ours is the only work to present any experimental results of computer algorithms for biometric recognition based on the ear.

- *K. Chang and K.W. Bowyer are with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556. E-mail: {kchang, kwb}@cse.nd.edu.*
- *S. Sarkar and B. Victor are with the Department of Computer Science and Engineering, University of South Florida, Tampa, FL 33620. E-mail: {sarkar, bvictor}@csee.usf.edu.*

Manuscript received 20 June 2002; revised 13 Dec. 2002; accepted 16 Feb. 2003. Recommended for acceptance by M. Pietikainen.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 116812.

2 "Eigen-Faces" and "Eigen-Ears"

Extensive work has been done on face recognition algorithms based on principal component analysis (PCA), popularly known as "eigenfaces" [5]. The FERET evaluation protocol [6] is the de facto standard in evaluation of face recognition algorithms, and currently uses PCA-based recognition performance as a baseline. A standard implementation of the PCA-based algorithm [7] is used in the experiments reported here. This implementation requires the location of two landmark points for image registration. For the face images, the landmark points are the centers of the eyes. Manually identified eye center coordinates are supplied with the face images in the Human ID database. For the ear images, the manually identified coordinates of the triangular fossa and the antitragus [1] are used. See Fig. 1 for an illustration of the landmark points.

The PCA-based approach begins with using a set of training images to create a "face space" or "ear space." First, the landmark points are identified and used to crop the image to a standard size located around the landmark points. In our experiments, original face images are cropped to $768 \times 1,024$ and original ear images to 400×500. In these images, one pixel covers essentially the same size area on the face or the ear. Next, the cropped images are normalized to the 130×150 size used by the PCA software. At this point, one pixel in an ear image represents a finer-grain metric area than in a face image. The normalized images are masked to "gray out" the background and leave only the face or ear, respectively. The face images use the mask that comes with the standard implementation [7]. For the ear images, we experimented with several different levels of masking in order to tune this algorithm parameter for good performance. Last, the image is histogram equalized. The eigenvalues and eigenvectors are computed for the set of training images, and a "face space" or "ear space" is selected based on the eigenvectors associated with the largest eigenvalues. Following the FERET approach, we use the eigenvectors corresponding to the first 60 percent of the large eigenvalues and drop the first eigenvector as it typically represents illumination variation [6]. This approach uses the same dimension of face space and ear space, 117, in this case (Table 1). Another approach is to use whatever number of eigenvectors accounts for some fixed percent of the total variation, resulting in a different dimension of face space and ear space. Which of these approaches is used does not substantially affect our conclusions, as is shown later in the paper.

The set of training images consists of data for 197 subjects, each of whom had both a face image and an ear image taken under the same conditions at the same image acquisition session. These images were acquired at the University of South Florida (USF) between August 2000 and November 2001. A subject's images were dropped from our study if either the face or ear was substantially obscured by hair, if the subject wore an earring or analogous face jewelry or if either image had technical problems. Some of the gallery and probe images for the first experiment were acquired at USF during the same time frame. Additional gallery and probe images for the first experiment, and all gallery and probe images for the second and third experiments, were acquired at the University of Notre Dame in November 2002.

There is a separate (gallery, probe) data set for each of three experiments. The gallery images represent the "watch list," that is,

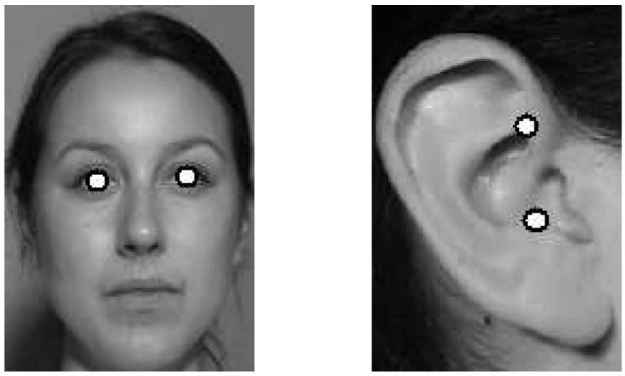

Fig. 1. Illustration of points used for geometric normalization of face and ear images. The triangular fossa is the upper point on the ear image and the antitragus is the lower point.

the people who are enrolled in the system to be recognized. A probe image is an image given to the system to be matched against the gallery. Each of the three experiments represents a single factor being varied in a consistent way between the gallery and probe. For the day variation experiment, 88 subjects had both an ear and a face image taken under the same conditions in one acquisition session and then another ear and face image taken under the same conditions on a different day. The face images are the standard FERET "F_A" ("normal expression") images [6]. The ear images are of the right ear. For each subject, the earlier image is used as the gallery image and the later image is used as the probe image. This experiment looks at the recognition rate when gallery and probe images of a subject are obtained on different days, but under similar conditions of pose and lighting.

For the lighting variation experiment, 111 subjects had an ear and a face image taken under the same conditions in one session and then another face and ear image taken in the same session, but under a different lighting condition. The standard lighting uses two side spotlights and one above-center spotlight and the altered lighting uses just the above-center spotlight. The images taken under the standard lighting are gallery images and the images taken under altered lighting are probe images. This experiment looks at the recognition rate when gallery and probe images of a subject are obtained in the same session and with similar pose, but under distinctly different lighting.

TABLE 1
A Number of Eigenvectors Used to Create the Eigenspace

Eigenvector Selections	Face	Ear	Face Plus Ear
First 60% of total eigenvectors	117	117	117
Eigenvectors used in 90% energy variation	86	76	102

For the pose variation experiment, 101 subjects had both an ear and a face image taken under the same conditions in one acquisition session and then another face and ear image taken at 22.5 degree rotation in the same acquisition session. The images taken from a straight-on view are the gallery set, and the images taken at a 22.5 degree rotation are the probe set. This experiment looks at the recognition rate when gallery and probe images of a subject are obtained in the same session and with the same lighting, but with a different pose. An example of the gallery and different probe conditions for one subject appear in Fig. 2.

Not all subjects attended all acquisition sessions and some subjects were dropped from some experiments after image quality control checks and, so, the three experiments have different numbers of subjects. The same standard face and ear images of some subjects may appear in the gallery set for each of the three experiments. However, since the probe sets are the changed conditions, there are no images in common across the three probe sets.

3 Experimental Results: Face versus Ear

The null hypothesis for these experiments is that there is no significant difference in performance between using the face or the ear as a biometric, given 1) use of the same PCA-based algorithm implementation, 2) the same subject pool represented in both the gallery and probe sets, and 3) controlled variation in one parameter of image acquisition between the gallery and probe images. The recognition experiment is to compute the cumulative match characteristic (CMC) curve for the gallery and probe set and to consider the statistical significance of the difference in rank-one recognition rates.

The baseline is the day variation experiment. This experiment looks at the recognition performance for gallery and probe images taken under the same conditions but on different days. The

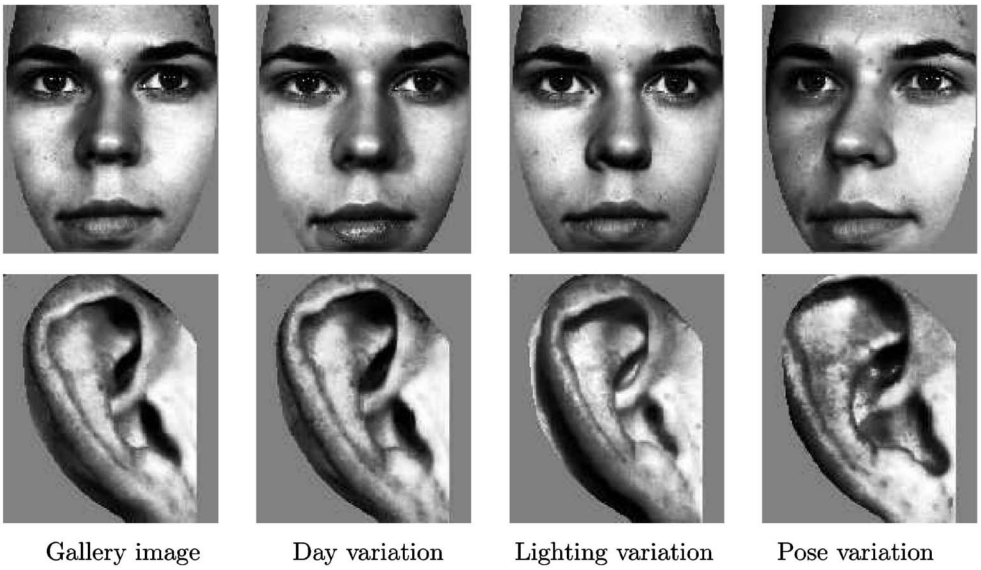

Fig. 2. An example of the gallery and probe face and ear images used in this study.

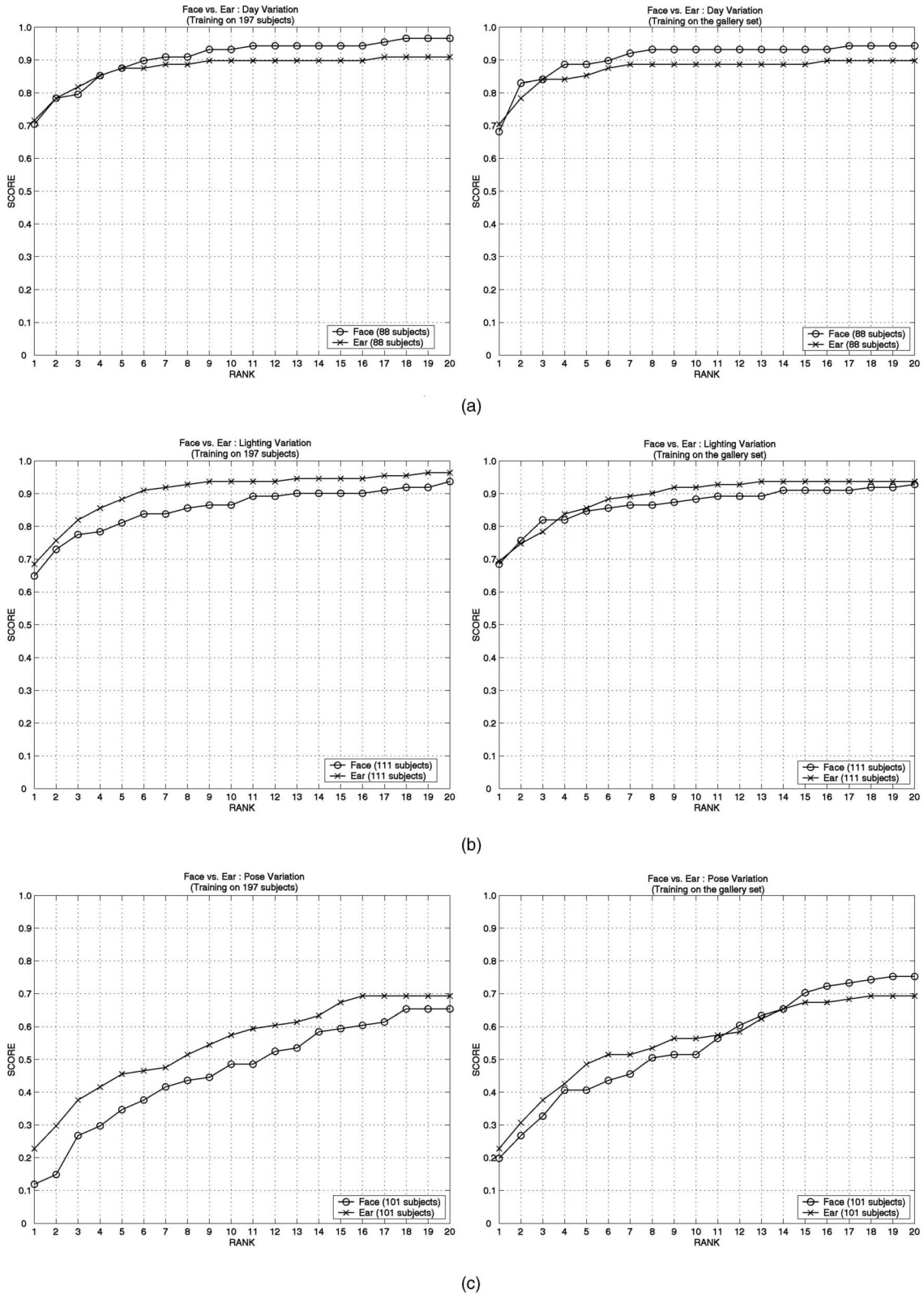

Fig. 3. Recognition performance comparison between face and ear. (a) Face and ear recognition performance in the *day variation* experiment. (b) Face and ear recognition performance in the *lighting variation* experiment. (c) Face and ear recognition performance in the *pose variation* experiment.

CMC curves for face and ear recognition are shown in Fig. 3. The CMC curves are computed in two ways. One uses the 197-image training set that has no subjects in common with the gallery and probe sets. The other uses the gallery set as the training set. There is no substantial difference in the results between the two training methods. Numbers reported for statistical significance tests are taken from the results using the 197-image training set. Note that the ear and face performance represented in the CMC curves is quite similar, with the curves actually crossing at some point. The rank-one recognition rates of 70.5 percent for face and 71.6 percent for ear

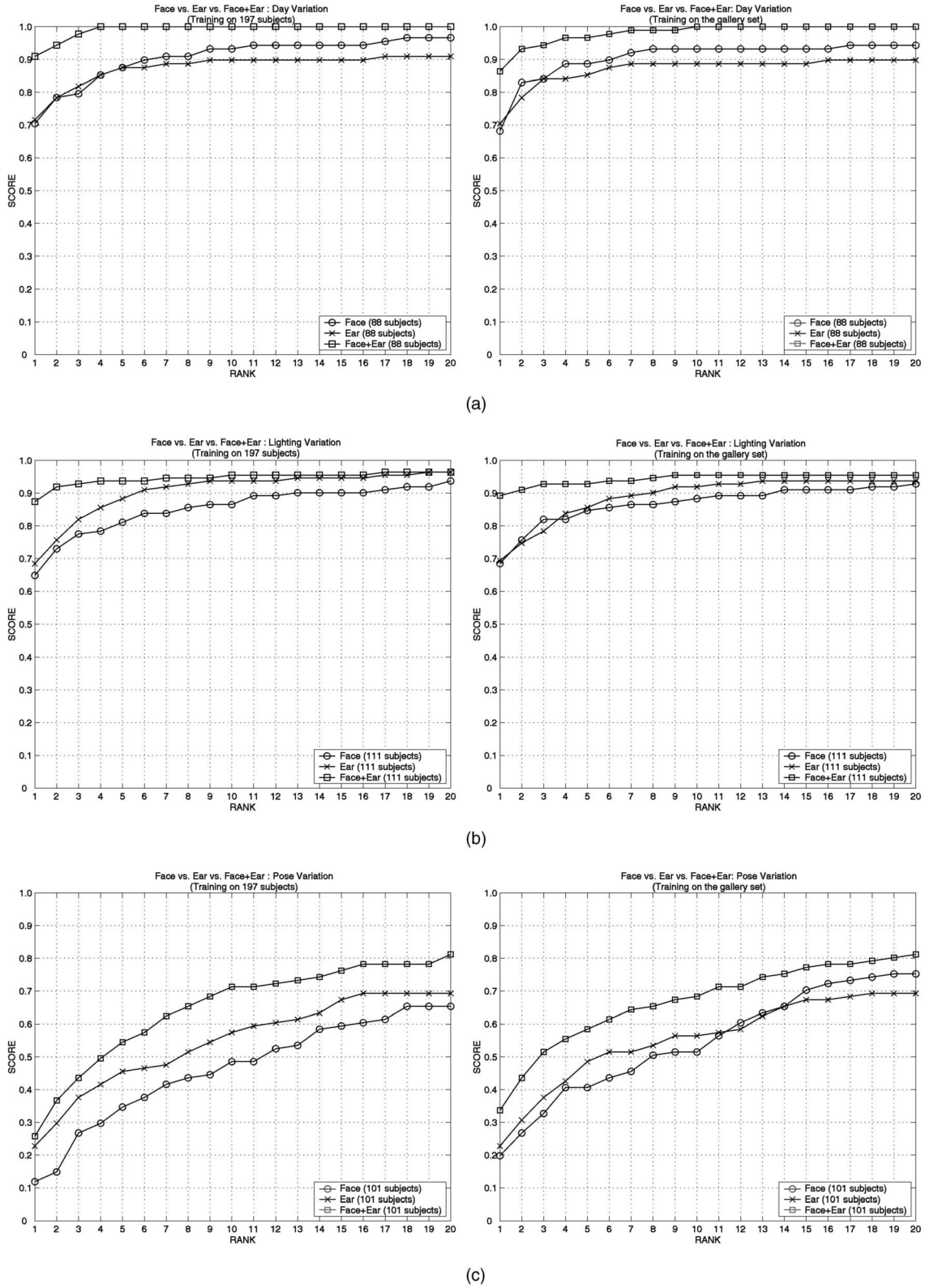

Fig. 4. Recognition performance of face, ear, and combined face-ear. (a) Face combined with ear recognition performance in the *day variation* experiment. (b) Face combined with ear recognition performance in the *lighting variation* experiment. (c) Face combined with ear recognition performance in the *pose variation* experiment.

are not statistically significantly different at the 0.05 level using a McNemar test [8].

Relative to the baseline experiment, the lighting variation experiment looks at how a lighting change between the gallery image and the probe image affects the recognition rate. Performance for either the face or the ear is slightly lower than in the baseline experiment. Similar to the baseline experiment, there is relatively little difference between the CMC curves for the face and the ear, especially at lower ranks. The rank-one recognition rates of

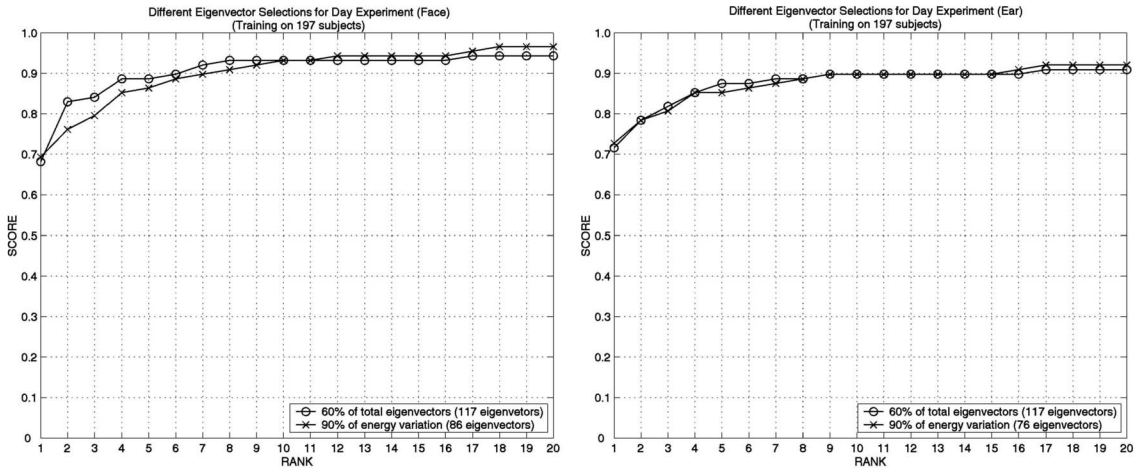

Fig. 5. Performance based on different selection of eigenvectors in face and ear spaces.

64.9 percent for face and 68.5 percent for ear are not statistically significantly different at the 0.05 level using a McNemar test.

Relative to the baseline experiment, the pose variation experiment looks at how a 22.5 degree rotation to the left between the gallery and the probe images affects the recognition rate. Performance, in this case, is much lower than for either the baseline or the lighting change experiment. There also appears to be a larger gap between face and ear performance than in the other two experiments, but still the difference is not statistically significant. In any case, performance at this low of a level is not likely to be practically meaningful.

Overall, the results of our experiments do not provide any significant evidence for rejecting the null hypothesis that the face and the ear have equal potential as the source for appearance-based biometric recognition. Of course, there may still be some biometric algorithm, other than PCA, for which one of the face or the ear offers significantly better recognition performance than the other. Also, there may be particular application scenarios in which it is not practical to acquire ear and face images that meet similar quality control conditions. For example, in an outdoor sports context many people may wear sunglasses or in a formal indoor event many people may wear earrings.

4 Experimental Results: Face Plus Ear Multimodal Biometric

Another experiment was performed to investigate the value of a multimodal biometric using the face and ear images. A very simple combination technique is used. The normalized, masked ear and face images of a subject are concatenated to form a combined face-plus-ear image. This was done with the data from each of the three experiments and Fig. 4 shows the resulting CMC curves. The CMC curves for the day variation and lighting variation experiments suggest that the multimodal biometric offers substantial performance gain. The difference in the rank-one recognition rates for the day variation experiment using the 197-image training sets is 90.9 percent for the multimodal biometric versus 71.6 percent for the ear and 70.5 percent for the face. A McNemar's test for significance of the difference in accuracy in the rank-one match between the multimodal biometric and either the ear or the face alone shows that multimodal performance is significantly greater at the 0.05 level. Of the 88 probes, the multimodal and the ear are correct on 62, both incorrect on 6, multimodal only is correct on 18, and ear only is correct on 2. The difference between the multimodal biometric and either the face or the ear alone is again statistically significant in the lighting change experiment, 87.4 percent rank-one recognition rate versus 64.9 percent or 68.5 percent, for the face or ear, respectively. However, because the overall performance is so low, the difference in the pose change experiment is not statistically significant. These results suggest that it is worthwhile to explore the combination of multiple biometric sources that could be acquired in a surveillance scenario.

5 Discussion

Overall, our experimental results suggest that the ear and the face may have similar value for biometric recognition. Our results do not support a conclusion that an ear-based or face-based biometric should necessarily offer better performance than the other. Of course, this is not the same as proving that there is no useful biometric algorithm for which one would offer better performance. Research into new algorithms that take advantage of specific features of the ear or the face may produce improved performance using one or the other.

Our results do support the conclusion that a multimodal biometric using both the ear and the face can out-perform a biometric using either one alone. There is substantial related work in multimodal biometrics. For example, Hong and Jain [9] used face and fingerprint in multimodal biometric identification, and Verlinde et al. [10] used face and voice. However, use of the face and ear in combination seems more relevant to surveillance applications. We are aware of just one other work specifically on multimodal biometrics appropriate to surveillance, this one using face and gait [11]. This would seem to be an especially rich and promising area of research. It might be expanded to include other biometric sources, such as face, ear, and gait. It might also be expanded to investigate more sophisticated methods of combining evidence from the different biometrics.

The results presented so far are based on using the same fixed number of eigenvectors for both the face and ear space. It is also possible to create the spaces based on the same percent of energy, allowing the number of eigenvectors to vary as appropriate. CMC curves computed using both spaces for the day variation experiment appear in Fig. 5. Performance is essentially the same whether the spaces are created based on a fixed number of eigenvectors in this case, or a floating number of eigenvectors corresponding to a fixed percent of total energy. (The authors would like to thank the anonymous reviewer who suggested inclusion of this comparison.)

The PCA-based face recognition approach has been informally tuned through use over time and, inevitably, an accumulation of expertise is embedded in the standard implementation [7]. Several options were explored in an attempt to ensure that the use of the PCA approach was appropriately tuned for use with ear images. For example, five different levels of masking for the ear images were tried. Also, a total of four landmark points were marked on each ear image and experiments were run with a different pair of landmark points. The results reported here are for the best level of masking and pair of landmark points.

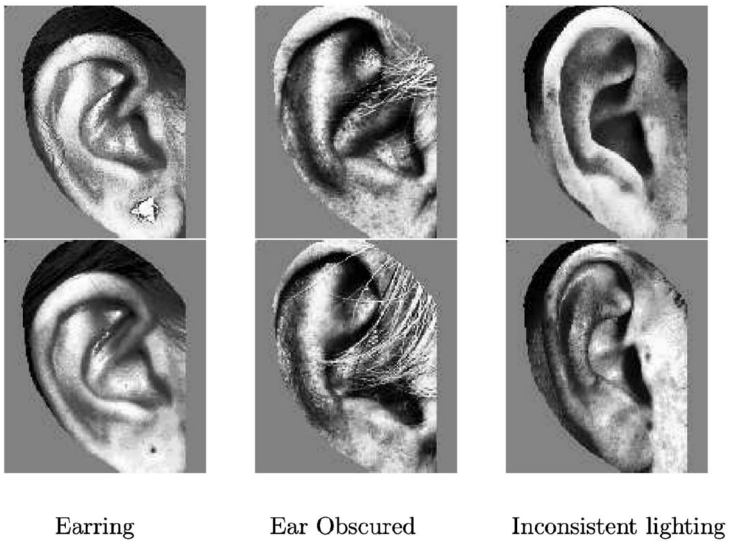

Fig. 6. Examples of (gallery, probe) image pairs not used in this study.

Our results are obtained using the PCA-based algorithm, whereas Burge and Burger [2] and Hurley et al. [3] each propose a different approach. Thus, one possible reservation to our conclusion is that it may be dependent on the particular algorithmic approach. However, we know of no experimental results in the literature for either of the other proposed approaches. In our own efforts to implement one of the approaches, we found the basic ear description used to be rather unstable. The description is an attributed graph obtained from the Voronoi diagram of the edges detected in the ear image [2]. One problem is that the edges detected from the ear image can be very different for relatively small changes in camera-to-ear orientation or in lighting. The edges detected in an image of the ear arise mostly from occluding contours, rather than from surface discontinuities or surface-marking-like effects. Thus, the edges will naturally be substantially different if there are changes in orientation or lighting.

We have tried to make the face versus ear aspect of this experiment "fair" in the sense of having equivalent quality control rules for each type of image. For example, all images are of subjects not wearing earrings or any face jewelry and all images had no substantial amount of the ear or face obscured by hair. These restrictions are in one sense equal in terms of quality of images used in the experiments, but are not necessarily equal in the sense of being equally likely to be true of images acquired in practice. For example, many more subjects were dropped from the experiments due to earrings, than due to face jewelry. Also, it may be more likely for hair to obscure the ear than the face. The question of whether the ear or face is more likely to be cleanly imaged, in practice, seems to depend on a number of cultural, social, and environmental factors, and is not dealt with in this study.

The results presented in this paper differ somewhat from those in the paper by Victor et al. [4]. Results of that study showed ear-based recognition performance was significantly lower than face-based performance. The image data sets in that study had less control over the covariates such as earrings, hair over ears, exact lighting setup over time, etc., and this variation in image quality confounded with the covariates under study. However, the results in that study might not be "wrong" so much as reflect the average quality of images likely to be acquired in real applications. Examples of images exhibiting such issues appear in Fig. 6. Even though only a small number of images in the previous study exhibited such quality control issues, these often resulted in misrecognition and, so, excluding them effectively increases the measured performance for the ear biometric.

The experimental materials used in this study are available to other researchers. The materials are distributed as a UNIX tar file containing the raw and masked images, the version of the PCA implementation used, and scripts that can be run to replicate the basic results. See www.nd.edu/~cvrl/ for information on obtaining the experimental materials.

ACKNOWLEDGMENTS

This work was supported by the DARPA "Human ID at a Distance" program, Air Force Office of Scientific Research contract F49620-00-1-0388, and Office of Naval Research contract N-000140210410. Thanks to P. Jonathon Phillips for his ideas and suggestions related to this work. Thanks to Laura Malave, Christine Kranenburg, Christine Bexley, Padmanabhan Soundararanjan, Earnie Hansley, and Isidro Robledo-Vega for help in preparing some of the data. The authors would also like to thank the anonymous reviewers of an earlier version of this paper for their consistent and on-target suggestions.

REFERENCES

[1] A. Iannarelli, *Ear Identification,* Forensic Identification Series. Fremont, Calif.: Paramont Publishing, 1989.
[2] M. Burge and W. Burger, "Ear Biometrics," *BIOMETRICS: Personal Identification in a Netwroked Society,* A. Jain, R. Bolle, and S. Pankanti, eds. pp. 273-286, Kluwer Academic Publishers, 1999.
[3] D.J. Hurley, M.S. Nixon, and J.N. Carter, "Force Field Energy Functionals for Image Feature Extraction," *Image and Vision Computing J.,* vol. 20, pp. 311-317, 2002.
[4] B. Victor, K.W. Bowyer, and S. Sarkar, "An Evaluation of Face and Ear Biometrics," *Proc. Int'l Conf. Pattern Recognition,* pp. 429-432, Aug. 2002.
[5] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.
[6] P.J. Phillips, H. Moon, S.Y. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Analysis and and Machine Intelligence,* vol. 22, no. 10, pp. 1090-1104, Oct. 2000.
[7] R. Beveridge and B. Draper, "Evaluation of Face Recognition Algorithms (release version 3.0)," http://www.cs.colostate.edu/evalfacerec/index.html, 2003.
[8] R. Beveridge, K. She, B. Draper, and G. Givens, "Parametric and Nonparametric Methods for the Statistical Evaluation of Human ID Algorithms," *Proc. Workshop Empirical Evaluation Methods in Computer Vision,* Dec. 2001.
[9] L. Hong and A. Jain, "Integrating Faces and Fingerprints for Personal Identification," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 12, pp. 1295-1307, Dec. 1998.
[10] P. Verlinde, G. Matre, and E. Mayoraz, "Decision Fusion Using a Multi-Linear Classifier," *Proc. Int'l Conf. Multisource-Multisensor Information Fusion,* vol. 1, pp. 47-53, July 1998.
[11] G. Shakhnarovich and T. Darrell, "On Probabilistic Combination of Face and Gait Cues for Identification," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 169-174, 2002.