Over Twenty Years of Eigenfaces

MATTHEW TURK, University of California, Santa Barbara

The inaugural ACM Multimedia Conference coincided with a surge of interest in computer vision technologies for detecting and recognizing people and their activities in images and video. Face recognition was the first of these topics to broadly engage the vision and multimedia research communities. The Eigenfaces approach was, deservedly or not, the method that captured much of the initial attention, and it continues to be taught and used as a benchmark over 20 years later. This article is a brief personal view of the genesis of Eigenfaces for face recognition and its relevance to the multimedia community.

Categories and Subject Descriptors: I.4.9 [**Image Processing and Computer Vision**]: Applications; I.5.4 [**Pattern Recognition**]: Applications; K.6.5 [**Management of Computing and Information Systems**]: Security and Protection

General Terms: Algorithms

Additional Key Words and Phrases: Face recognition, biometrics, vision-based interaction

ACM Reference Format:

Turk, M. 2013. Over twenty years of Eigenfaces. ACM Trans. Multimedia Comput. Commun. Appl. 9, 1s, Article 45 (October 2013), 5 pages.
DOI: http://dx.doi.org/10.1145/2490824

In the 20 years since its inaugural ACM conference in 1993, the multimedia research community has advanced the state of the art in technologies and systems that utilize images and video. During this time, the focus of this research has shifted from image-based signals to semantics, that is, from the effective management and display of image and video data to the understanding of content in order to enable a range of capabilities and applications. This transition has been fueled by progress in algorithms and systems from the field of computer vision, which has developed during this period from an area with few commercially viable applications 20 years ago to a field producing working solutions in a wide range of application areas. One of the most significant areas of overlap for the multimedia community is the processing and analysis of humans in images and video. Technologies for human face detection, face recognition, facial expression analysis, body tracking, gesture recognition, and activity analysis have emerged in recent years and show great promise for widespread use in practical multimedia and interactive systems for consumers, businesses, and governments.

In the years leading up to the 1st ACM International Conference on Multimedia, there was significant interest in face recognition for biometrics and surveillance applications, beginning with Bledsoe's report on *man-machine facial recognition* in 1966 and continuing with early influential work

Author's address: M. Turk, Department of Computer Science, University of California, Santa Barbara, CA 93106-5110; email: mturk@cs.ucsb.edu.

DOI: http://dx.doi.org/10.1145/2490824

by Kelly [1970], Kanade [1973], and Harmon et al. [1981]. By the late 1980s, there were two main directions of work in automatic face recognition: template-based (or appearance-based or holistic) approaches (e.g., Burt's pyramid machine [Burt 1988], Fleming and Cottrell's neural network approach [1990]) and feature-based approaches (e.g., Yuille et al.'s deformable templates [1989]). With a surge of interest in connectionism, there was also the question of whether the core learning and recognition architecture for face recognition should be a traditional classification approach (e.g., [Wong et al. 1989]) or a neural network approach (e.g., [Midorikawa 1988]). At the same time, there was a flurry of activity in visual neurophysiology and psychology on how humans (and other animals) represent and process faces, from low-level (face-selective neurons) to high-level (e.g., prosopagnosia, face inversion studies) phenomena. Two of the key questions of interest in both computational and biological communities were: (1) Is face recognition a special visual ability, somewhat independent of other recognition tasks? (2) What are the best ways to model and process faces to support recognition?

Intrigued by this multidisciplinary topic, Sandy Pentland and I pondered these issues for some time, seeking insight from both the computational vision and biological vision communities, but ultimately looking for a practical solution to a problem posed by a company (Arbitron, at the time seeking to compete with Nielsen in the television ratings business[1]) looking to incorporate face recognition into a set-top box for automatic TV ratings measurements. In order to recognize who is in the living room watching television, the system needed to efficiently and accurately detect, locate, and recognize faces among a relatively small population—family members, relative, friends—and to easily register new home viewers while ignoring the family dog on the couch. Unlike some other recognition problems, this implied mostly frontal view faces and scenarios where the participants would be mostly stationary for some time, with little concern for intentional deception by users. Although this differed in many respects from the typical biometrics scenario that drove much of the work in face recognition at the time, it fit well some consumer uses of face recognition that have become relevant to the multimedia community, such as processing faces in photographs and movies. Our primary constraints were a reasonable frame rate (recognition every two seconds was adequate and also ambitious at the time) and relatively robust performance.

Most of the work in recognition through the 1980s focused on feature detection, description, and matching, whether low-level or shape-based features. The connectionism resurgence had brought about recognition architectures that focused on learning, but left the intermediate representations and computations difficult to analyze. We began looking for a method that would steer away from explicitly describing and measuring isolated features and their relationships, taking the opposite strategy of holistic, appearance-based representations; we also wanted to keep clear of the "black box" approach of the neural networks approach in order to have a better ability to understand and debug the method.

Following the scientific tradition of standing on the shoulders of giants (and perhaps also on the long tradition in computer vision of re-inventing things done previously in the photogrammetry and pattern recognition communities), we came across the 1987 paper by Sirovich and Kirby [1987] that used dimensionality reduction (specifically principal components analysis) to represent face images. There had been prior work in pattern recognition on using statistical techniques to reduce dimensionality for classification problems (e.g., [Fukunaga 1990]), but Sirovich and Kirby's insight was to apply this specifically to representing frontal face images, which could be aligned to a configuration shared by all members of the population. With PCA as the core representation mechanism, we could build an approach to face recognition that used a holistic, image-based representation while retaining the ability to understand the intermediate representations: the Eigenfaces. Conceptually, these Eigenfaces

[1]Interestingly, as of this writing, Nielsen is in the process of acquiring Arbitron.

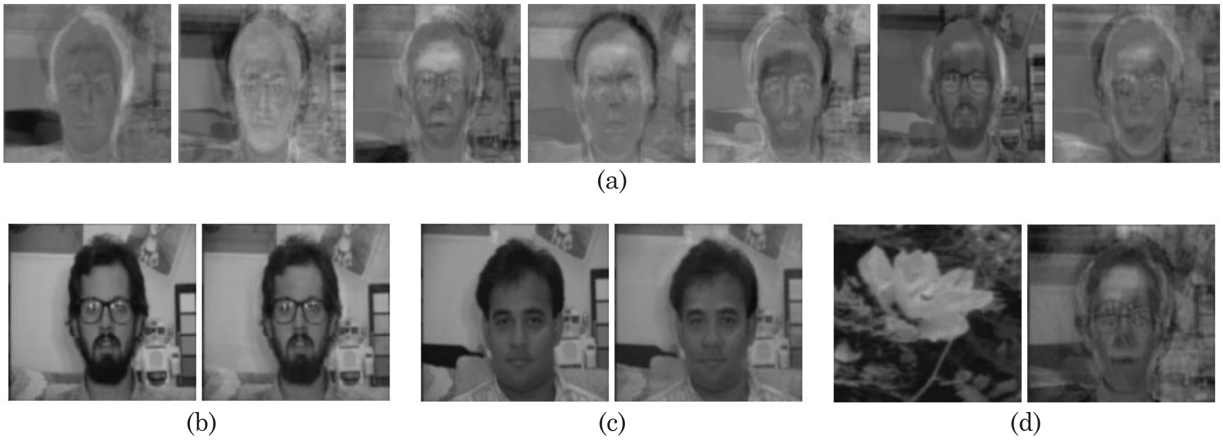

Fig. 1. (a) The first seven Eigenfaces from a small databases of face images. (In this early example from Turk [1991], the backgrounds were included in the computation; the background was later masked and not included in the computation.) (b)–(d) Original images (left in each pair) and their projections into face space (right). The Euclidian difference between left and right images is the distance from face space; the larger the distance, the less likely it is that the original image is of a face.

span the subspace of all possible face images, or the *face space*; image deviations (whether due to image noise or other factors, such as illumination, pose, expression, occlusions, etc.) push an image away from the space, and the *distance from face space* can be used to determine how likely an image is to be a face in the first place, thus providing a built-in mechanism for face detection (see Figure 1).

Despite the attention it received, it was clear from the start that the Eigenfaces approach to face recognition was limited. It was constrained to frontal views of faces; it could not distinguish between changes in imaging conditions, face pose, expression, occlusions, and identity. The linear subspace representation was restrictive, as was the initial Euclidian face space measure used to determine identity. Although multiview and multiscale representations could provide some degree of pose and scale independence, "in between" views and scales were problematic. The ability of the system to scale to thousands or millions of face images was questionable at best. Most concerning, perhaps, was the inability to predict and thus compute parameters needed for real systems, such as the number and variety of training images, and the number of eigenfaces required, for a given scenario.

Nevertheless, for a variety of reasons, this initial Eigenfaces work caught the attention of many computer vision researchers—the timing was right for the pendulum to swing in the other direction for a while, toward image-based approaches in recognition. My dissertation [Turk 1991] included examples of eigeneyes and eigenexpressions, and soon there were other eigen-recognition papers in vision, audio (e.g., eigenvoices [Kuhn et al. 1998]), and other domains. Pentland et al. [1994] and Moghaddam et al. [1998] and others worked towards generalizing the approach to handle a wider range of conditions. For better or worse, Eigenfaces became the default comparison method for much of the work in face recognition for the next decade or two, and graduate (and undergraduate) students continue to learn the approach in introductory vision courses. Over 20 years later, I still get occasional emails from students around the world who have questions about or want to discuss the approach.

Has all this attention been warranted? The debate continues, but if the popularity of Eigenfaces has helped to inspire researchers to create better face recognition algorithms and approaches and to provide a useful example of an image-based approach to recognition, then it has served a useful purpose. It is, however, time to retire Eigenfaces as a modern approach to face recognition and as a baseline comparison method, as it has long ago been eclipsed by more robust methods. “Better performance than Eigenfaces” is no longer a meaningful metric.

My personal experience with this work and its aftermath leads me to offer some unsolicited advice for young researchers.

—Explore your interests, especially relevant interdisciplinary interests, outside your core topic or department. They may lead you in fruitful directions you would not have otherwise anticipated.
—Don’t just wait for the great, out-of-the-blue idea to build your research on; explore the breadth and depth of the landscape in your area of interest, and be on the lookout for prior work that may be appropriate for a new context.
—Be opportunistic regarding building on the work of others (while being sure to generously acknowledge such work). This is how it should work; you get a better view when standing on the shoulders of giants.
—If you become known for a specific research area or contribution (often your dissertation work), continue to follow your interests, whether that continues in the same direction or not. Although I’m still viewed by many as the “Eigenfaces guy,” I subsequently pursued other interesting research directions and have only occasionally done research on face recognition. Don’t let others shoehorn you into a single area.

Back to the historical narrative for a moment: since these early days of automatic face recognition, the ACM Multimedia community has both pushed the state of the art of face processing and been a consumer of face recognition technologies for a range of multimedia application areas. Several topics of current interest to the community, such as multimedia content-based indexing and retrieval, visual search, human-centric media, multimedia and multimodal interaction, meeting/presentation analysis, event recognition, and social media, depend on face detection and recognition capabilities. This synergy between the computer vision and multimedia communities is beneficial to both. As the two communities continue to mature, I look forward to continued synergy and great progress in technologies and applications of mutual interest.

REFERENCES

BURT, P. J. 1988. Smart sensing within a pyramid vision machine. *Proc. IEEE 76*, 8, 1006–1015.

FLEMING, M. AND COTTRELL, G. 1990. Categorization of faces using unsupervised feature extraction. In *Proceedings of the International Joint Conference on Neural Networks*. Vol. 2, 65–70.

FUKUNAGA, K. 1990. *Introduction to Statistical Pattern Recognition*. Academic Press.

HARMON, L. D., KHAN, M. K., LASCH, R., AND RAMIG, P. F. 1981. Machine identification of human faces. *Pattern Recognition 13*, 2, 97–110.

KANADE, T. 1973. Picture processing system by computer complex and recognition of human faces. Ph.D. dissertation, Department of Information Science, Kyoto University.

KELLY, M. D. 1970. Visual identification of people by computer. Stanford Artificial Intelligence Project Memo AI-130.

KUHN, R., NGUYEN, P., JUNQUA, J.-C., GOLDWASSER, L., NIEDZIELSKI, N., FINCKE, S., FIELD, K., AND CONTOLINI, M. 1998. Eigenvoices for speaker adaptation. In *Proceedings of the International Conference on Spoken Language Processing*.

MIDORIKAWA, H. 1988. The face pattern identification by back-propagation learning procedure. In *Abstracts of the 1st Annual INNS Meeting*. 515.

MOGHADDAM, B., WAHID, W., AND PENTLAND, A. 1998. Beyond eigenfaces: Probabilistic matching for face recognition. In *Proceedings of the 3rd International Conference on Automatic Face- and Gesture-Recognition*. 30–35.

PENTLAND, A., MOGHADDAM, B., AND STARNER, T. 1994. View-based and modular eigenspaces for face recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*. 84–91.

SIROVICH, L. AND KIRBY, M. 1987. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A* 4, 3, 519–524.

TURK, M. 1991. Interactive-time vision: Face recognition as a visual behavior. Ph.D. dissertation, MIT Media Lab, Cambridge, MA.

WONG, K., LAW, H., AND TSANG, P. 1989. A system for recognising human faces. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1638–1642.

YUILLE, A. L., COHEN, D. S., AND HALLINAN, P. W. 1989. Feature extraction from faces using deformable templates. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*.

Received May 2013; accepted May 2013