# Imagery as a Biologically Motivated Enhancement to PCA-Based Face Matching*

Marcio L. Teixeira and Bruce A. Draper

Computer Science Department, Colorado State University, Fort Collins, CO 80523,
USA.

**Abstract.** Prior work on face recognition has focused on the use of dimensionality reducing transformations, such as PCA or LDA, to reduce the complexity of the incoming data. These algorithms perform matching in a compressed space, typically choosing the nearest neighbor as the match. We propose that while such transformations are useful for choosing a set of candidates for matching, the selection among the candidates is best performed in uncompressed image space. This paper describes a biologically inspired technique that augments reduced dimension matching with post processing in the original uncompressed space. Our system, which uses a simple perceptron for the post processing stage, shows an improvement over standard techniques.

## 1 Introduction

Depictive theories of mental imagery suggest that the brain can recreate highly accurate and image-like representations of visual memories. These theories, which were originally proposed to explain psychological data, have recently been bolstered by PET and rTMS data that show that mental imagery activates area 17 in the primary visual cortex – the same area used by the early vision system [10]. In addition, some researchers believe that the ability to recreate virtual images from memories evolved to assist visual recognition as well as imagery [9].

PET and fMRI studies have also given us new insights into how people recognize faces. Face recognition, like all visual tasks, activates the early vision system in the visual cortex. In addition, however, creates foci of activity in the fusiform gyrus [17, 3, 6] and the right inferior frontal gyrus, an area previously associated through lesion studies with visual memory [12] (see also [14]). The implication is that faces are recognized by preprocessing the retinal images in the visual cortex, and then matching them in the fusiform gyrus to memories of other faces stored in the inferior frontal gyrus. Taken together with the mental imagery studies, we can further suggest that, when needed, the visual memories are sufficient to recreate eidetic images in the visual cortex.

This paper is motivated by the question of whether eidetic memory reconstruction is beneficial to face recognition, particularly if we assume a subspace

---

* The work was funded in part by the Defense Advanced Research Projects Agency (DARPA) under contract DABT63-00-1-0007.

projection and matching model for face recognition. In particular, we consider the hypothesis that fine discriminations between images and memories may be easier in image space than in a compressed subspace. To test this idea, we implemented a two-stage model in which test images are projected into a subspace and compared to a gallery of compressed memories. (This is a standard technique for automatic face recognition.) The system does not, however, match the image to a memory in subspace. Instead, it selects the N closest memories in subspace, and creates N difference images (in uncompressed image space) between the test image and the N selected memories. A simple perceptron is then used to select the difference image that is most likely to have been generated by two images of the same subject. Our two-stage system combines the best attributes of systems that perform matching entirely in a subspace with those that perform matching in entirely in image space. Our system provides the accuracy of the latter with the efficiency of the former.

## 2 Prior work

Early evidence for mental image reconstruction came from experiments that asked subjects to remember and answer questions about previously viewed scenes; the results suggested that subjects reconstructed and then mentally scanned the image [8]. Priming studies provided convergent psychological evidence: visual priming is stronger when subjects are asked to imagine an object rather than repeat a sentence about it ([9], Chapter 9); images of incomplete objects prime for images of the missing parts [13]; and the priming effects from imagined masks are as strong as those from observed masks [4]. The most compelling evidence, however, comes from brain imaging studies: PET studies show that mental imaging tasks activate the same area in the primary visual cortex (Area 17) as visual recognition does, while rTMS studies show that disabling Area 17 prevents both recognition and mental imaging [10]. It should be noted, however, that not all researchers agree with the depictive theory of visual imagery (for a dissenting opinion, see [18]).

The pathway associated with face recognition was first identified in fMRI studies in the mid-1990s [17, 3, 6]. These studies scanned patients while they viewed images of human faces. The scans revealed activation not only in the primary visual cortex, but also in the fusiform gyrus. Subsequent PET studies imaged a larger portion of the brain and confirmed the activation in the fusiform gyrus, while adding another locus of activity in the right inferior frontal gyrus, an area previously associated through lesion studies with visual memory [12] (see also [14]). Moreover, in both fMRI and PET studies, the activation was specific to recognizing faces and other familiar objects [2, 5, 21, 22].

In recent years, work on automatic (machine) face recognition has been dominated by subspace projection techniques. Kirby and Sirovich first showed that principal component analysis (PCA) is optimal in the sense that it minimizes the mean squared error between the original images and their reconstructions for any given level of compression [7]. Turk and Pentland were the first to apply PCA to face or object recognition [23]. They used PCA to compute a set of subspace basis vectors (which they termed "eigenfaces") for a database of face images, and projected the images onto those vectors. New test images were matched to images in the database by projecting them onto the basis vectors and finding the nearest image in the subspace.

The initial success of eigenfaces popularized the idea of matching images in compressed subspaces and lead to a search for other subspaces. Other methods of generating subspaces for face recognition were proposed, including linear discriminant analysis [20], independent component analysis (ICA [1]), non-negative matrix factorization (NMF [11]), and Gabor jets [15]. With the exception of LDA, all of the subspace techniques satisfy the biological model in the sense that they (1) project images into a compressed subspace for subsequent matching to previously memorized images, and (2) are invertible, so that images can be reconstructed from their compressed representations, as is required for mental imagery. (LDA projections are not invertible.) PCA is at one end of the spectrum, in that it is optimal for reconstruction, but not necessarily for matching. Not surprisingly, subspace projection techniques have been suggested as models of human face and object recognition (e.g. [24]).

An interesting variation on subspace projection matching was put forth by Moghaddam and Pentland [16]. As with the work cited above, they used PCA to project their data into a subspace. What is unusual is that they projected difference images, rather than raw images, into the subspace. They cast the multi-class problem of distinguishing among images of different subjects into a dual class problem of distinguishing among intrapersonal and extrapersonal difference images. Our system shares this approach.

## 3 Intrapersonal and Extrapersonal Subspaces

In traditional classifiers, face images are projected directly into a compressed subspace, under the assumption that images of a single person will map to a tight cluster of points. Conversely, it is expected that the projections of images of different subjects will be widely separated. Certain subspace methods, such as LDA, make this objective explicit by purposefully trying to choose a projection that minimizes the interclass separation while maximizing the in-between class separation [20]. Such methods, however, assume that projections exist that map facial images onto non-overlapping regions for each subject (nearest-neighbor classifiers, for instance, depend heavily on this).

Moghaddam and Pentland propose an alternative. Their classifier defines the subspace in a different way: rather than treating face images as points in a face subspace, they instead look at the space spanned by the *difference* between two face images. The difference image for two face images is the signed arithmetic difference between respective pixels in the source images. Such difference images fall into two distinct classes: *intrapersonal* difference images are those derived from two images of the same subject, while *extrapersonal* difference images are derived from two images of different subjects. Moghaddam and Pentland suggest that intrapersonal and intrapersonal difference images form distributions that are approximately Gaussian [16]. Their classifier matches probe images to stored images by computing the likelihood that the corresponding difference images came from the subspace of interpersonal rather than extrapersonal images.

### 3.1 Face Recognition with Image Feedback

Subspace matching is a possible model for face recognition, because we believe that faces are matched to memories in the fusiform gyrus, and that these memories are sufficient to form eidetic image reconstructions in the primary visual cortex. Therefore visual memories are a form of compressed image and subspace matching becomes the obvious model for biological face recognition.

We consider the hypothesis that fine discrimination in image matching may be easier in image space than in a compressed subspace. To test this hypothesis, we created a two-stage model in which probe images are projected into a PCA subspace and compared to compressed gallery images (i.e. memories). Our model does not match the images in subspace, however. Instead, it selects the N closest gallery images and then creates N difference images in the original image space. A perceptron is then used to select the difference image that is most likely to have been generated by two images of the same person.

The value of N is a tunable parameter in this model, and the extreme cases are degenerate: when N is one, the system behaves as a nearest-neighbor subspace classifier and the difference images do not affect the outcome. When N is the number of gallery images, distance in subspace becomes irrelevant and classification is done entirely based on the difference images. Interestingly, our experiments show best results for small values of N.

The assumption underlying this model is that image matches are most easily discriminated by looking at the spatial distribution of the errors, which is not explicit in a compressed subspace. We use PCA to generate subspaces because it is the most thoroughly tested and understood of the subspace projection techniques; ICA, NMF or Gabor filters could have been used just as easily. We use a nearest neighbor algorithm to match images in subspace for similar reasons. It is interesting that the resulting system uses similar components to those used by Moghaddam and Pentland, but in the opposite order. Moghaddam and Pentland create all possible difference images, and then project the difference images into subspace. We project raw images into subspace, select a subset of the nearest gallery images, and then create the difference images only for that limited subset.

### 3.2 Results

Figure one presents our results for the four standard FERET test sets, as defined in [19]. In each case, we use traditional PCA-based matching to gather a set of "nearest neighbor" gallery images for each probe image. For each of these gallery images, we generate the signed difference image of the probe and the gallery image. The difference image is downsized to 16x16 image and presented to a 256 input perceptron for classification. The network is trained on half of the original probe images and tested on the other half. The training set was itself was partitioned into two parts, one half used for adjusting the network weights the other half for verification (to avoid over-fitting to the training data).

**Performance Graphs** Our results are presented in Fig. 1. We vary the number of nearest gallery images that were chosen for each probe image and denote this number N. The horizontal axis of the graph shows N on a log scale: it ranges from one to the size of the entire gallery (1196 images). The vertical axis of the graph shows the fraction of images correctly identified. The plot compares our two-stage classifier to two baseline classification strategies: *nearest-neighbor* and *cumulative N-rank* matching. The nearest neighbor strategy always picks the gallery image which is nearest the probe in PCA subspace. Its performance is a straight line that coincides with the performance of our two-stage classifier when N is one. At the other extreme, the cumulative N-rank classifier examines the N closest images in subspace, and is considered to have recognized the probe image correctly if *any* of the N closest gallery images are correct. The N-rank classifer cannot be implemented in practice because it relies on *a priori* knowledge that is not available to a real system. However, it is interesting because it illustrates the upper bound on the performance of our two-stage matcher: the two-stage matcher cannot succeed if none of the N nearest neighbors are correct.

We find that for small values of N, the performance for the two-stage classifier is better than the nearest neighbor classifier. The improvement is particularly significant for the fb, fc and dup2 data sets. Notice that the performance of the two-stage matcher quickly reaches a plateau, suggesting that there is little benefit to constructing all the difference images: choosing a small value of N not only leads to better accuracy but significantly reduces the computational costs.

Overall, our experiments with the two-stage matcher suggest that the spatial distribution of errors is useful for making fine discriminations among closely matching images, but it cannot match two widely disparate images of the same subject. Interestingly, this agrees with Kosslyn's conjecture that memory reprojection is used only in ambiguous cases [9].

**Perceptron Weights** To shed light on how the perceptron operates, Fig. 2 shows the relationship between the network weights and the average of the first 40 FERET face images. The first row shows the average face for each probe set, the second row shows locations where the network weights are above a positive threshold, and the third row shows locations where the network weights are below a negative threshold. As the perceptron sees difference images, rather than faces, the figures show how the perceptron chooses to weight disparities lying in different locations of the image: those falling in the positive areas favor an increase in the networks output, while those falling in negative areas favor a decrease the network output.

Observing the weights yields some general insights. In particular, we see that the network is more likely to trigger a mismatch when the errors are located in the positive regions near the eyes, forehead and nose. Errors falling in the

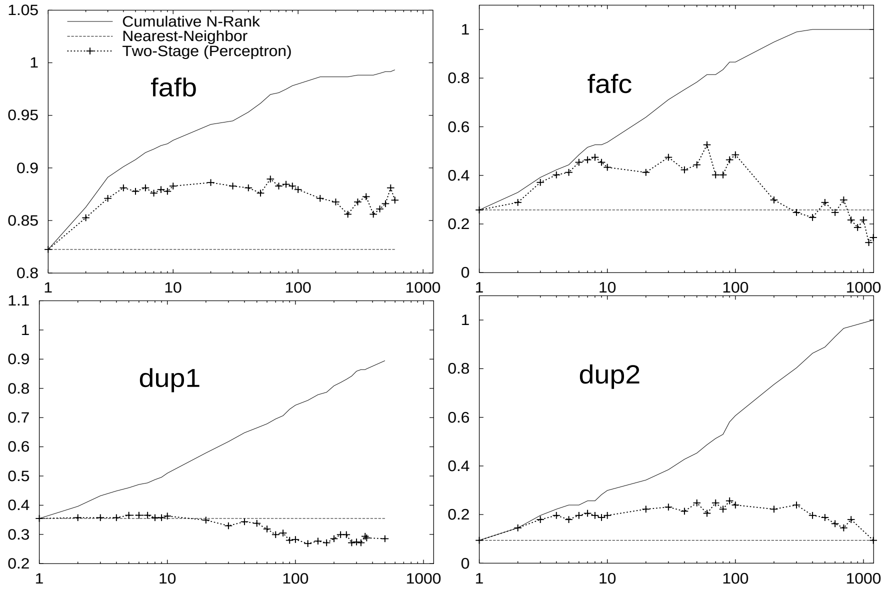

**Fig. 1.** These graphs show performance on the fafb, fafc, dup1 and dup2 datasets. Due to prohibitive computational costs, the fafb and dup1 plots were not computed to the full extent of N.

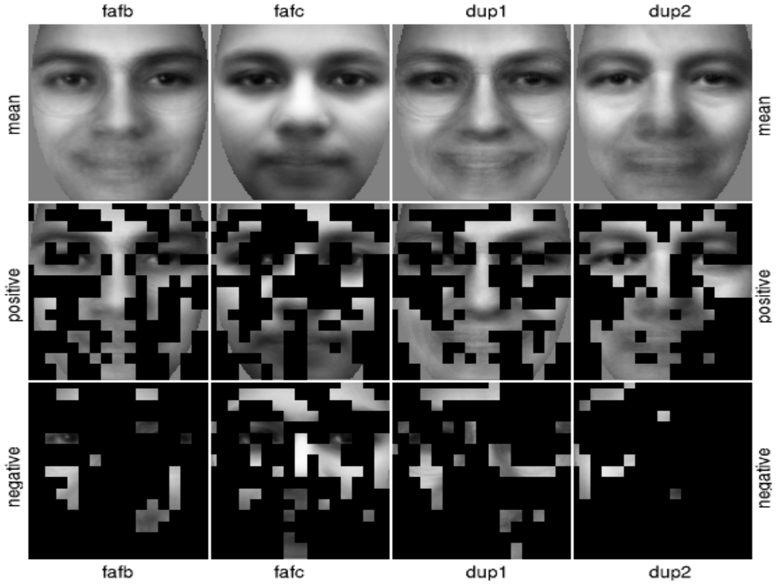

**Fig. 2.** Mean image and areas corresponding to positive and negative perceptron weights. Only weights whose absolute value exceed 15% of the total range are shown.

negative regions – predominantly the cheek and mouth – actually increase the possibility of a match. We conjecture that errors around the cheek often result from specular highlights, so the perceptron learns to disregard them. A similar argument may be made for the mouth, which is sensitive to changes in expression. The weights suggest that the perceptron learns to focus on specific areas on the image and in so doing is able to improve on the baseline matching provided in subspace.

## 4 Conclusions and Future Work

We are able to improve performance over traditional subspace matching by using a perceptron to classify the difference images generated from the probe image and the few nearest gallery images in subspace. The network weights suggest that the perceptron learns to exploit spatial properties of the difference images that are not explicit in the compressed subspace. Recognition performance peaks when ten or so candidates are chosen for each probe. If you go beyond this, the performance begins to drop, suggesting that generating more spatial difference images would increase the computational workload without any gains in performance. These results are consistent with a biological model of face recognition as subspace projection matching, with eidetic memory reconstruction for fine discrimination in ambiguous cases.

There is still much work to be done. We cannot yet explain the system's poor performance on the dup1 probe set. We need to explore the effects of using difference images reconstructed from the PCA eigenspace, rather than relying on the difference of the original images, as this would more accurately model the biological system. We also plan to compare the performance of our approach to the Bayesian classifier by Moghaddam and Pentland.

## References

1. M.S. Bartlett. *Face Image Analysis by Unsupervised Learning,* Kluwer Academic, Boston. 2001.
2. L. L. Chao, A. Martin, and J. V. Haxby. "Are face-responsive regions selective only for faces?," *NeuroReport*, 10:2945-2950, 1999.
3. V. P. Clark, K. Keil, J. M. Maisog, S. Courtney, L. G. Ungeleider, and J. V. Haxby. "Functional Magnetic Resonance Imaging of Human Visual Cortex during Face Matching: A Comparison with Positron Emission Tomography," *NeuroImage*, 4:1-15, 1996.
4. A. Ishai and D. Sagi. "Common Mechanisms of Visual Imagery and Perception," *Science*, 268:1772-1774 (June 23, 1995).
5. A. Ishai, L. G. Ungerleider, A. Martin, J. L. Schouten, and J. V. Haxby. "Distributed representation of objects in the human ventral visual pathway," *Science*, 96:9379-9384, 1999.
6. N. Kanwisher, M. Chun, J. McDermott, and P. Ledden. "Functional Imaging of Human Visual Recognition," *Cognitive Brain Research*, 5:55-67, 1996.

7. M. Kirby and L. Sirovich. "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103-107, 1990.
8. S.M. Kosslyn and S.P. Shwartz. "A Simulation of Visual Imagery," *Cognitive Science*, 1:265-295, 1977.
9. S.M. Kosslyn. *Image and Brain: The Resolution of the Imagery Debate*, MIT Press: 1996
10. S.M. Kosslyn, A. Pascual-Leone, O. Felician, S. Camposano, J.P. Keenan, W.L. Thompson, G. Ganis, K.E. Sukel and N.M. Alpert. "The Role of Area 17 in Visual Imagery: Convergent Evidence from PET and rTMS," *Science*, 284:167-170 (April 2, 1999).
11. D. Lee and S. Seung. "Learning the Parts of Objects by Non-negative Matrix Factorization," *Nature* 401:788-791.
12. E. Maguire, C. D. Frith, and L. Cipolotti. "Distinct Neural Systems for the Encoding and Recognition of Topography and Faces," *NeuroImage*, 13:743-750, 2001
13. P. Michelon and O. Koenig. "Imagery Feedback in Object Identification," Object Perception and Memory Conference, 1999.
14. K. Nakamura, R. Kawashima, N. Sata, A. Nakamura, M. Sugiura, T. Kato, K. Hatano, K. Ito, H. Fukuda, T. Schormann, and K. Zilles. "Functional delineation of the human occipito-temporal areas related to face and scene processing: a PET study," *Brain*, 123:1903-1912, 2000.
15. K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. von der Malsburg, "The Bochum/USC Face Recognition System And How it Fared in the FERET Phase III Test," *Face Recognition: From Theory to Applications*, H. Wechsler, P.J. Phillips, V. Bruce, F. Fogelman Soulie, T.S. Huang (Eds.). Springer-Verlag, pp.186-205, 1998.
16. B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-19(7):696-710, July 1997
17. A. Puce, T. Allison, J. C. Gore, and G. McCarthy. "Face-sensitive regions in human extrastriate cortex studied by functional MRI," *Journal of Neurophysiology*, 74:1192-1199, 1995.
18. Z.W. Pylyshyn. "Mental Imagery: In search of a theory," *Behavioral and Brain Sciences*, in press.
19. P.Rauss P.J.Philips, H.Moon and S.A.Razvi. The feret evaluation methodology for face recognition algorithms. IEEE Conference on Computer Vision and Pattern Recognition.
20. D. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 18(8):831-836, 1996.
21. M. J. Tarr and I. Gauthier. "FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise," *Neuroscience*, 3:764-769, 2000.
22. F. Tong, K. Nakayama, M. Moscovitch, O. Weinrib, and N. Kanwisher. "Response Properties of the Human Fusiform Face Area," *Cognitive Neuropsychology*, 17:257-279, 2000.
23. M. Turk and A. Pentland, "Eigenfaces for Recognition." *Journal of Cognitive Neuroscience*, 3(1):71-86, 1991.
24. S. Ullman. "Three-dimensional Object Recognition Based on the Combination of Views," *Object Recognition in Man, Monkey, and Machine,* M. Tarr and H. Bulthoff (eds), MIT Press, Cambridge MA, 1998.