Can a Linear Autoassociator Recognize Faces From New Orientations?

Dominique Valentin and Hervé Abdi

School of Human Development, The University of Texas at Dallas, Richardson, TX 75083-0688, and Université de Bourgogne à Dijon, Dijon 21004 France.

Received: February 1995, accepted: June 1995, revised version received: October 1995

An often noted limitation of computational models of faces operating on 2D pixel intensity representations is that they cannot handle changes in orientation. We show that this limitation can be overcome by using multiple views of a given face instead of a single view to represent the face. Specifically, we show that a linear autoassociator trained to reconstruct multiple views of a set of faces is able to recognize the faces from new view angles. An analysis of the internal representation of the memory (*i.e.,* eigenvectors of the between unit connection weight matrix) shows a dissociation between two kinds of perceptual information: orientation *versus* identity information.

1. Introduction

In recent years, quite a number of computational/statistical models of face processing have been proposed in the literature. These models have been successfully applied to a wide range of tasks such as image compression, face detection, categorization, recognition, and identification, as well as feature detection and selection (see [1, 2] for reviews). Much of the effort going into these recent models has been concentrated on the processing of single frontal (or nearly frontal) two-dimensional (or 2D, in brief) pixel-based representations of faces. Consequently, their performance is sensitive to substantial variations in lighting conditions, size, and position in the image. To avoid these problems, a preprocessing of the faces is necessary. However, this can be done in a relatively straightforward manner by using automatic algorithms for locating the faces in the images and normalizing them for size, lighting and position (*e.g.,* [3]).

Most models in the literature assume that the preprocessing step has already been implemented and focus on the problem of face recognition *per se.* An additional and more drastic limitation of pixel-based representations is that they are not three-dimensionally (3D) invariant, and hence the performance of models operating on this type of representation across large changes in orientation is rather poor. The problem, therefore, is to know if it is possible to extend existing models so that they can handle depth rotation.

The methodology of this study is based on empirical evidence suggesting that human subjects trained with multiple 2D views of unfamiliar faces handle depth rotation better than human subjects trained with single views of the same faces [4, 5]. This could be due to several reasons. First, subjects could elaborate a 3D invariant representation of the faces, such as the object-centered structural description described by Marr and Nishihara [6] for object recognition. However, although the object-centered approach seems relevant for the general domain of object recognition, its extension to the specific case of face recognition is problematic. While object recognition requires stimuli to be assigned to broad categories that maximize the physical similarities between exemplars (*i.e.,* the "basic level categories" as defined by Rosch [7]), face recognition requires the discrimination of stimuli *within* the basic level category "face." Object-centered models seem more appropriate for assigning objects to basic level categories than for discriminating instances within basic level categories.

Some recent studies in computational theory of 3D object recognition suggest that, for basic level categories, objects might be more efficiently represented using a set of viewer-centered representations than a single object-centered representation [8, 9, 10]. These representations depend on the position of the viewer relative to the object to be recognized, and therefore are specific to the particular viewpoint from which the object is perceived. Applied to the problem of face recognition, a face could be represented in memory by a limited set of 2D view-dependent descriptions that correspond to its familiar orientations. Recognition could be achieved by transforming an input representation of a face to the orientation of the nearest stored representation [11], or, if the faces are represented by a series of views that are close enough to each other, by interpolating among those views [12, 13, 14].

In the series of simulations we report here, we trained an autoassociative memory using multiple views of a set of faces, and we tested its ability to generalize to new views of the learned faces. Autoassociative memories are a powerful tool for storing, recognizing, and categorizing faces represented as pixel-intensity images [15, 16, 17, 18]. Their power stems mainly from the fact that they 1) use extremely efficient and well explored computational algorithms (*i.e.,* eigenvalue and singular value decomposition) and 2) are easily analyzable in terms of traditional mathematical concepts and statistical techniques (*i.e.,* least-squares estimation and principal components analysis). Because of these properties, autoassociative memories are also a useful tool for compressing [19, 3] and analyzing [20] the perceptual information in faces.

In the framework of image compression, the principal components analysis (PCA) technique (or Karhunen-Loève transform or singular value decomposition, cf. [21]) amounts to computing the eigenvectors of the pixel-covariance matrix of a set of images represented as gray-level vectors. The eigenvectors constitute an orthogonal basis for representing (*i.e.,* in the eigenspace) the images. The images can be either perfectly represented using the complete eigenspace, or estimated by using a least-squares low-dimensional representation made of the eigenvectors with the largest eigenvalues. Since an object is represented as a point in a multidimensional space, its distance from the other objects in the space can be computed and used as a basis for the recognition process (*e.g.,* nearest neighbor algorithm). This approach as been applied recently to the problems of 3D object recognition by Murase and Nayar [22] and 3D face recognition by Pentland, Moghaddam, and Starner [23].

Murase and Nayar represented a set of objects using two different types of eigenspaces. A *universal eigenspace*, and an *object eigenspace.* The universal eigenspace is used to identify the object presented as prompt. Once the object is identified as a particular object, it is projected onto the appropriate object eigenspace and its orientation is determined.

A somewhat different approach was used by Pentland *et al.* [23]. Instead of the universal and object eigenspaces proposed by Murase and Nayar, they used a multiple set of *view-based eigenspaces.* Each eigenspace is obtained by computing the eigenvectors of an image set of different individuals in a common orientation. The orientation of a face is first determined by computing its distance from each separate eigenspace. The projection coefficients of the face onto the closest eigenspace are then used to "identify" the face using a nearest-neighbor algorithm.

The autoassociative model we present here is somewhat similar to the approach of both Murase and Nayar [22] and Pentland *et al.* [23] in that using an autoassociative memory to store and retrieve faces is equivalent to computing the eigendecomposition of the set of faces and representing the faces as a weighted sum of eigenvectors [15]. Specifically, since the autoassociative memory is trained with multiple views of a set of faces, it is equivalent to the *universal eigenspace* representation proposed by Murase and Nayar, with the exception that in our application, all the eigenvectors of the set of faces are kept. The rationale for keeping the complete eigenspace is that we are more interested in using the eigenvector representation as a tool for analyzing the perceptual information in faces than for compressing the information. In the second part of the series of simulations reported below, we show that an advantage of having a representation in the complete universal eigenspace is that both the information about the identity and the orientation of the faces are preserved.

This paper is organized as follows: after presenting a brief description of a face autoassociative memory, we report a series of simulations showing that, when multiple 2D views of faces are used to train an autoassociative memory, then the memory can deal with the problem of depth rotation. Next, we analyze the perceptual properties of the internal representation developed by the autoassociative memory (*i.e.,* the eigenvectors of the weight matrix).

2. Linear autoassociator

A linear autoassociator (or autoassociative memory) is a network of simple units interconnected by a set of weighted connections. Learning occurs by modifying the weight of the connections. Autoassociative memories are content-addressable because, when part of a stimulus is given as a memory key, the memory gives back the complete stimulus, filling in the missing components.

To store a face in an autoassociative memory, the face is first digitized as a pixel image and the rows of the image concatenated to form an $I \times 1$ pixel vector $\mathbf{x}_k$ (where I is the number of pixels used to represent a face, and k indicates a particular face). Each numerical element in $\mathbf{x}_k$ is the gray level of the corresponding pixel. For computational convenience, the vectors $\mathbf{x}_k$ are normalized so that $\mathbf{x}_k^T\mathbf{x}_k$=1. The set of training faces is represented by an $I \times K$ matrix $\mathbf{X}$ obtained by concatenating the pixel vectors (where K is the number of training faces). The faces are stored in the memory by setting the weights of the connections between cells. These values are stored in an $I \times I$ matrix $\mathbf{W}$ using Hebbian learning. The Hebbian learning rule in a linear associator sets the change of the connection weights to be proportional to the product of the input and the output. For an autoassociator, the changes are proportional to the product of the input vector by its transpose (*i.e.,* the "outer product" of the vector and itself). With a proportionality constant equal to 1, $\mathbf{W}$ is obtained as:

$$(1) \qquad \mathbf{W} = \mathbf{X}\mathbf{X}^T = \sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^T$$

where T denotes the transpose operation.

Retrieval of a face is done by presenting the face as input to the memory. Specifically, recall of the k-th face is achieved as

$$(2) \qquad \widehat{\mathbf{x}}_k = \mathbf{W}\mathbf{x}_k$$

where $\widehat{\mathbf{x}}_k$ represents the answer of the memory. The quality of this answer can be estimated, either by visually comparing the reconstructed face with the original face or, more formally, by computing the cosine of the angle between $\mathbf{x}_k$ and $\widehat{\mathbf{x}}_k$:

$$(3) \qquad \cos(\widehat{\mathbf{x}}_k, \mathbf{x}_k) = \frac{\widehat{\mathbf{x}}_k^T \mathbf{x}_k}{||\widehat{\mathbf{x}}_k||\ ||\mathbf{x}_k||}$$

where $||\mathbf{x}_k||$ is the Euclidean norm of vector $\mathbf{x}_k$ (*i.e.,* $||\mathbf{x}_k|| = \sqrt{\mathbf{x}_k^T\mathbf{x}_k}$). A cosine of 1 indicates a perfect reconstruction of the stimulus.

Figure 2 (middle row) displays the response of a memory trained with Hebbian learning when a learned face, a new face, and a random pattern are presented as input. Clearly, the memory gives the same response for every stimulus, and hence it is not able to discriminate between faces or even to discriminate a face from a random pattern. The performance of the memory can be improved by using the Widrow-Hoff error-correction learning rule. This is equivalent to using a gradient descent method to adjust the weights of the connections so as to reduce the squared error between the faces and their reconstructions. The values of the weights are first computed using Hebbian learning and then iteratively corrected using the error signal (*i.e.,* the difference between the input face and the answer of the memory). Specifically, the weight matrix is expressed at time $n+1$ as

$$(4) \qquad \mathbf{W}_{[n+1]} = \mathbf{W}_{[n]} + \eta(\mathbf{X} - \mathbf{X}\mathbf{W})\mathbf{X}^T$$

where n represents the iteration number and η is a small positive constant (called the *learning constant*). Note that, as we shall detail later (cf. Eq. 6), the Widrow-Hoff learning rule can be implemented more simply by using the eigendecomposition of $\mathbf{W}$ [24, 25]. After complete Widrow-Hoff learning, all the faces in the training set are perfectly reconstructed, and if new faces are used as memory cues, they are

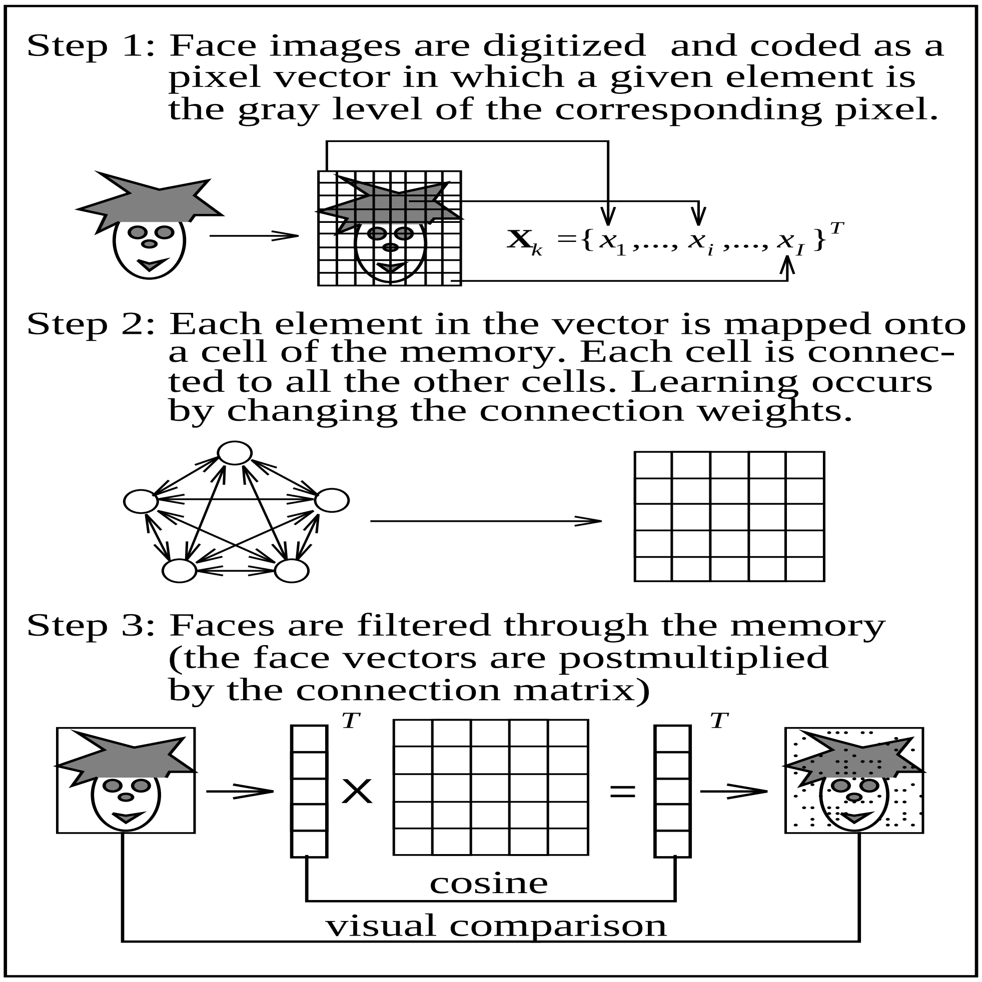

FIGURE 1. Illustration of a face autoassociative memory

distorted as an inverse function of their similarity with the training faces (cf. Figure 2, bottom row).

Previous studies (see [26] for a review) showed that autoassociative memories constitute a useful tool for recognizing and categorizing faces. Yet, the problem with these previous studies is that, because they used single views of the faces (generally a frontal view) as training sets, memory performance was very sensitive to depth rotation. The first objective of the simulations we present in the next section was to determine whether this limitation of autoassociative memories is a definite flaw, or whether it can be overcome by using multiple views as training sets.

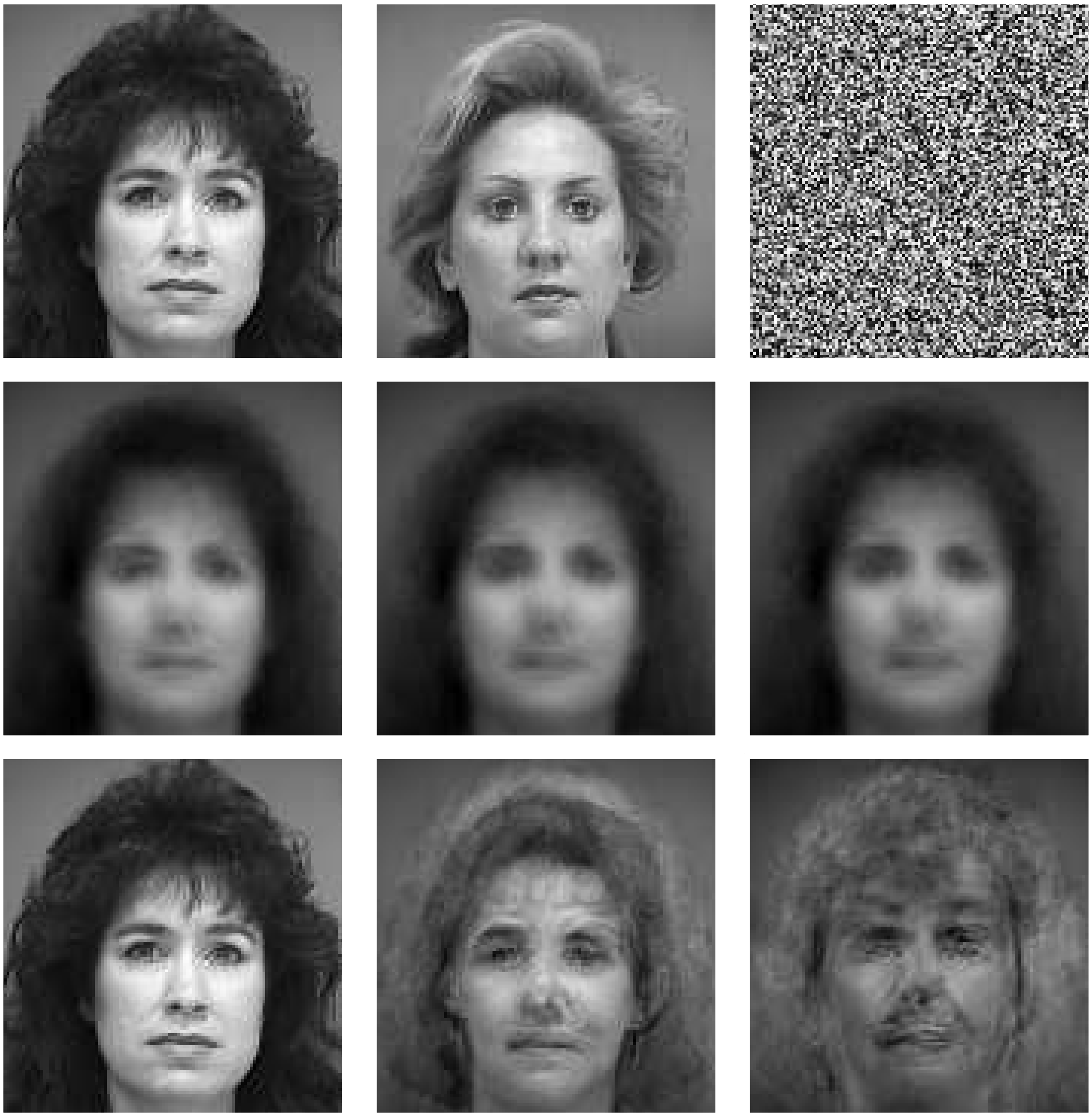

FIGURE 2. The top row shows three stimuli, the middle row shows the responses produced by an autoassociative memory trained with Hebbian learning to these stimuli, and the bottom row shows the responses produced by an autoassociative memory trained with Widrow-Hoff learning to the same stimuli. The stimuli are from left to right 1) a learned face, 2) a new face, and 3) a random pattern. When Hebbian learning is used, the memory produces the same response for all stimuli. Note, however, that the squared correlation between this answer and the face stimuli ($r^2 = .82$ for the learned face, and $r^2 = .42$ for the new face) is larger than that between the answer and the random pattern ($r^2 = 0$). When Widrow-Hoff learning is used, the learned face is perfectly reconstructed ($r^2 = 1$), and the new stimuli are distorted as an inverse function of their similarity with the learned face ($r^2 = .62$ for the new face, and $r^2 = 0$ for the random pattern).

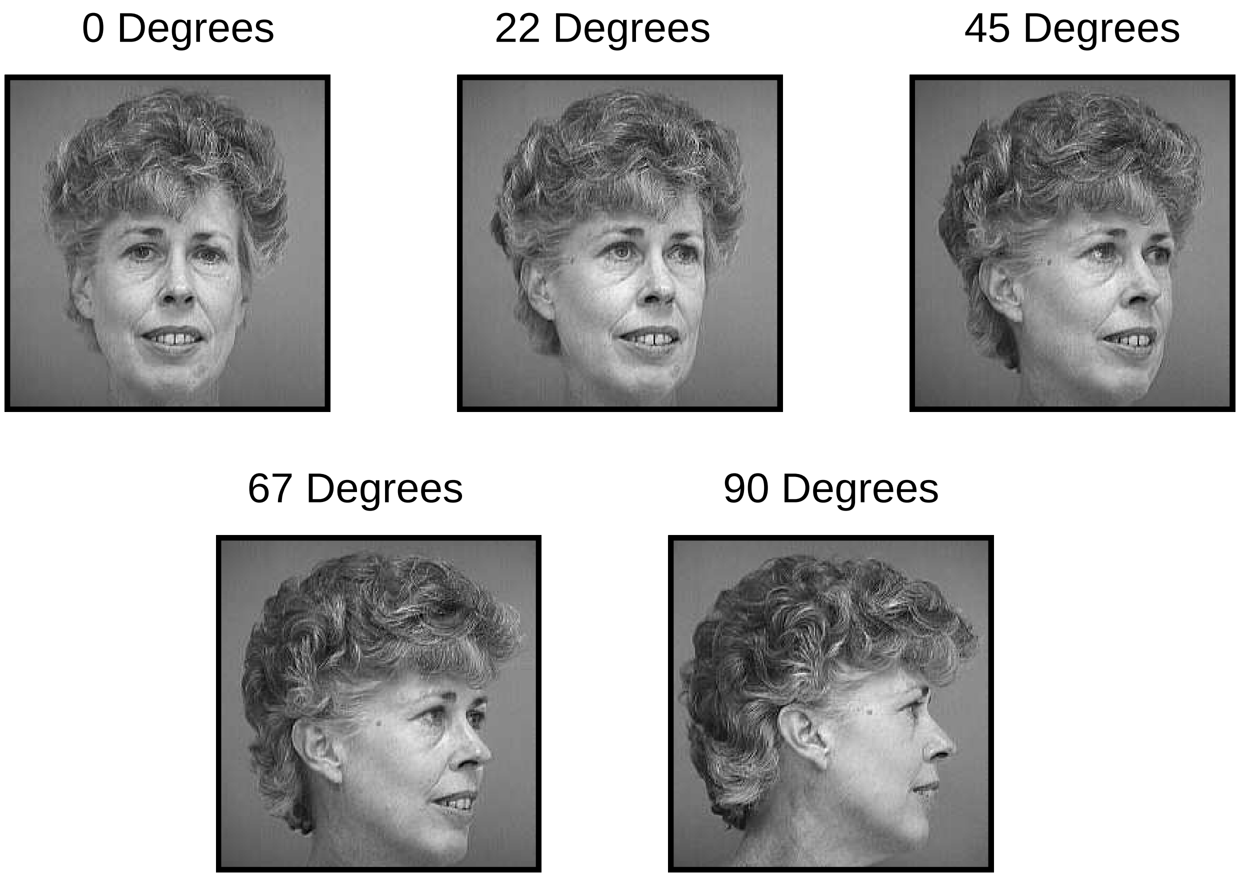

FIGURE 3. Examples of 5 stimuli used in the 4-views learning condition. The head samples the rotation in depth with about 20 degree steps.

3. Simulations

The purposes of the series of simulations presented in this section are:

- To test the ability of an autoassociative memory to recognize faces from new orientations when different numbers of views of the faces are used as training sets.
- To analyze the perceptual properties of the internal representation elaborated by a memory trained on multiple views of faces.

There are several ways to test the ability of an autoassociative memory to recognize faces. An easy one is to determine if the memory can distinguish between *learned* and *new* faces. This can be done by training the memory to reconstruct a set of target face images. After completion of learning, a new set of face images, composed of an equal number of learned faces and new faces (or distractors), is presented as input to the memory. For each face, target or distractor, the quality of the answer of the memory is estimated (*e.g.,* by computing the cosine between input and output as described by equation 3). If, on the average, the quality of reconstruction of target faces exceeds the quality of reconstruction of distractor faces, the model is said to distinguish between learned and unlearned faces. (*i.e.,* the model "*recognizes*" the old faces).

This procedure was used in the first simulations to analyze the ability of the memory to generalize the information learned from a limited set of views of faces to new views of the faces.

3.1. Recognition task.

3.1.1. *Stimuli.* Thirty female faces were used as stimuli. The faces were represented either by 10 views sampling the rotation of the head from full-face to profile in about 10-degree steps, or by 5 views sampling the rotation of the head in about 20-degree steps (cf. Figure 3). The face images were first digitized with a resolution of 230×240 pixels with 256 gray levels *per* pixel and then compressed by local averaging using a 2×2 window to give 115×120 pixel images. The images were roughly aligned along the axis of the eyes so that the eyes of all faces were about the same height. None of the pictured faces had major distinguishing characteristics such as clothes, jewelry, or glasses.

3.1.2. *Procedure.* The procedure included two phases: a learning phase, and a testing phase in which we tested the ability of the memory to "recognize" learned faces presented from new view angles. Half the faces were used as *targets* and the other half were used as *distractors.* During the learning phase, the target faces were stored in an autoassociative memory using complete Widrow-Hoff learning. The faces were represented using either: a single view (1-view condition), 4 views (4-view condition), or 9 views (9-view condition), sampling the rotation of the head from profile to full-face. In all learning conditions, 15 faces were used as training faces. In the 1-view and 4-view conditions, 1 or 4 views, respectively, of each training face were randomly chosen from a set of 5 possible views (0, 22, 45, 66, 90 degrees from full-face) to form the training set. In the 9-view condition, 9 views of each training face were randomly chosen from a set of 10 possible views (0, 10, 20, 30, 40, 50, 60, 70, 80, 90 degrees from full-face) to form the training set. In all conditions, the remaining views were used as testing views. After learning, all the views of the target were perfectly reconstructed (*i.e.,* the cosines between the original faces and their reconstructions were equal to 1).

During the testing phase, *new* views of the target faces and all the views of the distractors were filtered through the memory. For each view, a cosine was computed between the original and reconstructed images (cf. Eq. 3). The cosine provides an indication of the "familiarity" of the model with the face. The higher the cosine is, the more probable it is that the face has been learned. The recognition task was implemented by setting a criterion cosine, and by categorizing every face with a cosine greater than the criterion as "learned" and every face with a cosine smaller than the criterion as "new". Different values of the cosine were used as the criterion to generate a receiver operating characteristic (ROC, cf. [27]) for each view of the faces.

This procedure was repeated until each view of the target faces was used, in turn, as the "new" view during testing. Then, all the previous target faces were exchanged with the distractors and the procedure repeated until each view of the *new* target faces was used as the "new" view during testing. This was done to increase the size of the testing set. All the results reported below were computed on a total of 30 faces, each tested in five view angles (0, 22, 45, 66, 90 degree rotation from frontal view) for each of the three learning conditions (1-view, 4-view, and 9-view conditions).

3.1.3. *Results and discussion.* Figure 4 represents the quality of reconstruction of the faces as a function of the view orientation for each learning condition.

Three major points can be noted from this figure:

- The greater the number of views used to represent the faces, the better the quality of reconstruction.
- All the views are not equally well reconstructed. Specifically, the frontal and profile views are less well reconstructed than the intermediary views.
- When only a single view of the faces is used to train the memory, there is no major difference in the quality of reconstruction between targets and distractors. When four views are used, the targets are better reconstructed than the distractors. The difference between targets and distractors increases with the number of views used to represent the faces.

In summary, it seems that using several views of the faces as a training set increases the general reconstruction performance of the memory, as well as its ability to discriminate between learned faces presented from a new view angle and new faces. Besides this first approximation, we also need to look at recognition performance of the memory as measured by the area under the ROC (also named "area under the curve"). This is the signal-detection measure used to assess the discriminability of the new stimuli.

The area under the curve provides an unbiased estimate of the proportion of correct classification where chance performance is .50 [27].

Figure 5 displays recognition performance averaged across view angles. Clearly, in the 1-view condition the hit rate is equivalent to the false-alarm rate for all values of the criterion, and therefore the memory is not able to "recognize" the faces on which it has been trained when those faces are presented in novel orientations (area under ROC=.51). In the 4-view condition, the performance of the memory is slightly better (hit rate systematically greater than false-alarm rate), but still not very impressive (area under ROC =.56). In contrast, in the 9-view condition, the memory is clearly able to discriminate between learned and new faces (area under ROC =.79).

This performance is somewhat poorer than the performance reported in previous work by Pentland *et al.* [23]. Using a set of nine "view-based" eigenspaces to estimate first the orientation of the face, and then recognize it, they obtained an average performance of 86% correct recognition. The model proposed by Pentland *et al.* differs in two ways from the autoassociative model described here. First, in Pentland *et al.*'s model, separate eigendecompositions are used for different orientations of the faces, whereas a single eigendecomposition ("universal eigenspace") is used in our model. While the "view-based" solution is likely to be more accurate than the "universal" solution, it has the disadvantages of 1) assuming that the orientation of the faces is known at the time the eigendecomposition is computed, and 2) being computationally intensive since it requires as many eigendecompositions as orientations. The advantages of the autoassociative model, besides its simplicity, is that 1) it does not assume any prior knowledge about the orientation of the faces, and 2) (as we shall see in the next section) it dissociates spontaneously the information relative to the orientation of the faces from the information relative to their identity.

The second difference between the Pentland *et al.* model and the autoassociative memory resides in the algorithm used to recognize the faces. Pentland *et al.* used a nearest neighbor algorithm to decide whether a face is recognized or not, whereas we used a signal detection method. In this framework, the nearest neighbor algorithm searches for the smallest Euclidean distance between the projection onto the eigenspace of the face image to be recognized, and the projection of the learned faces. If this distance is smaller than a certain threshold, the face is recognized, and classified as its nearest neighbor. If the smallest distance is above the threshold, then it is characterized as unknown.

The main difference between the signal detection and the nearest neighbor methods is that the nearest neighbor keeps an explicit record of the faces to be classified, whereas the signal detection method is based *only* on the responses of the memory. While we can expect these two algorithms to yield roughly comparable performance when using the same information, further work is needed to assess the comparative merits of these methods quantitatively.

Figure 6 displays the recognition performance of the memory when 1) a profile, 2) a 3/4, and 3) a frontal view of the faces was presented as input for the testing phase. Again, it appears that all the views of the faces are not equally easy to recognize. While recognition performance is always greater in the nine-views condition than in the two other view conditions for all view angles, it is for the 3/4 view that this difference becomes very clear. In other words, it seems that 3/4 views are easier to recognize than either the profile or frontal views.

In summary, the results reported in this section indicate that when enough views of the faces are used as training sets, an autoassociative memory is able to handle the problem of depth rotation.

3.2. **Eigenvector Representation.** One advantage of using an autoassociative memory to store faces is that, since the weight matrix $\mathbf{W}$ is a cross-product matrix (and therefore is positive semi-definite), it can be analyzed in terms of its eigendecomposition [28, 16] as

$$\mathbf{W} = \mathbf{X}\mathbf{X}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad \text{with} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I} \tag{5}$$

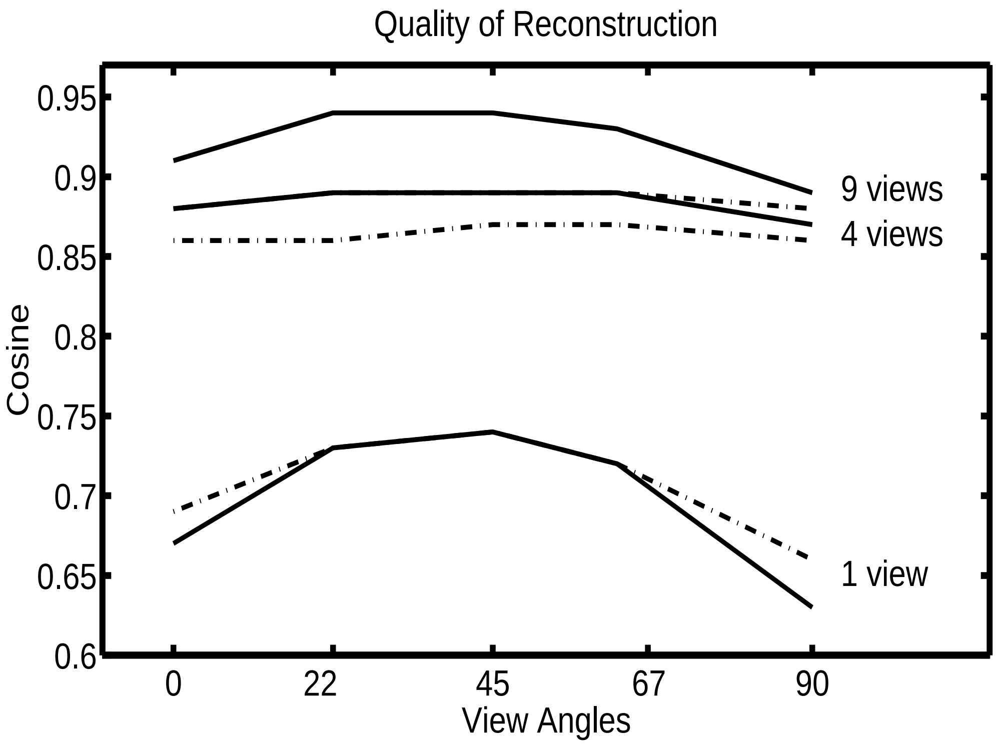

FIGURE 4. Quality of reconstruction of the targets and distractors as a function of learning condition. Targets are represented with a solid line and distractors with a dashed line.

where $\mathbf{U}$ is the matrix of eigenvectors of $\mathbf{W}$, and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. The eigendecomposition of $\mathbf{W}$ can be used also to analyze the Widrow-Hoff learning rule. Specifically, Equation 4 can be rewritten as (cf. [24])

$$\mathbf{W}_{[n+1]} = \mathbf{U}\mathbf{\Phi}_{[n]}\mathbf{U}^T \tag{6}$$

with

$$\mathbf{\Phi}_{[n]} = \mathbf{I} - (\mathbf{I} - \eta\mathbf{\Lambda})^n \tag{7}$$

which shows that Widrow-Hoff learning affects only the eigenvalues of $\mathbf{W}$. In particular, if η is properly chosen ([25]), $\Phi_{[n]}$ converges toward the unity matrix, and $\mathbf{W}_{[n]}$ from Eq. (6) converges towards

$$\mathbf{W}_{[\infty]} = \mathbf{U}\mathbf{U}^T \tag{8}$$

which is equivalent to saying that the weight matrix is sphericized [24] (*i.e.,* all its eigenvalues are equal to one).

Along the same lines, retrieval of a given face by the memory can be expressed as a weighted linear combination of the eigenvectors of $\mathbf{W}$. This is shown by combining Eqs. 2 and 5 as

$$\hat{\mathbf{x}}_k = \mathbf{W}\mathbf{x}_k = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\mathbf{x}_k = \sum_{\ell=1}^{L} \lambda_\ell \mathbf{u}_\ell \mathbf{u}_\ell^T \mathbf{x}_k \ , \tag{9}$$

where the scalar $\mathbf{u}_\ell^T\mathbf{x}_k$ corresponds to the projection of the kth face onto the ℓth eigenvector.

If complete Widrow-Hoff learning is used (cf. Eq. 8), Eq. 9 reduces to

$$\hat{\mathbf{x}}_k = \sum_{\ell=1}^{L} \mathbf{u}_\ell \mathbf{u}_\ell^T \mathbf{x}_k \ , \tag{10}$$

Interestingly, since the weight matrix is a *pixel* cross-product matrix, its eigenvectors can be graphically displayed and visually analyzed. Previous studies using this type of analysis showed that:

- The eigenvectors of a face autoassociative memory are face-like and can be interpreted as some kinds of macrofeatures or building blocks from which the faces are made [15, 19].
- A face can be approximated using a small number of eigenvectors [19, 3].

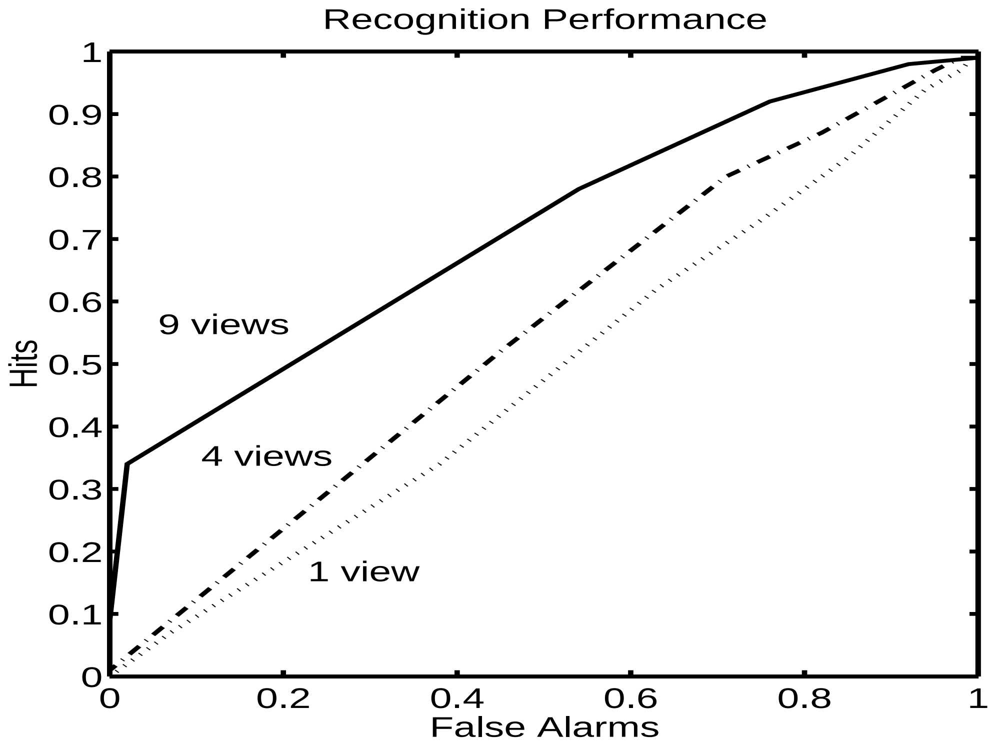

FIGURE 5. ROC curve as a function of learning conditions.

- Different ranges of eigenvectors convey different types of information. Specifically, eigenvectors with large eigenvalues convey mostly information useful for categorizing the faces along general semantic dimensions such as sex or race. Conversely, eigenvectors with low eigenvalues convey information useful for identifying specific faces [21, 20].

Using this eigendecomposition technique, we examined the perceptual information conveyed by different ranges of eigenvectors extracted from a cross-product face matrix built from 5 views of 30 female faces. As indicated previously, the five views sampled the rotation of the head in approximately 20-degree steps from full-face to profile (cf. Figure 3). By visually examining faces reconstructed using either the first ten eigenvectors (the ones with the largest eigenvalues), a middle range of eigenvectors (11 to 50), or the last 100 eigenvectors (the ones with the smallest eigenvalues), and by looking at the projections of the faces onto individual eigenvectors, we found that when multiple views of faces are used as input, the eigenvectors with large eigenvalues capture information relative to the orientation and general shape of the faces. These eigenvectors are useful in detecting the particular pose of the faces. In contrast, eigenvectors with intermediate eigenvalues contain information specific to small sets of faces across orientations. Eigenvectors with the smallest eigenvalues are specific to particular faces in particular orientations (or in a limited range of orientations). The dissociation between orientation and identity information is illustrated in Figure 7.

An additional examination of individual eigenvectors showed that for our sample of faces, the first eigenvector represents some kind of average across faces and poses. The second and third eigenvectors oppose profile views to frontal views for all the faces. The projections of a face onto these eigenvectors could be used as a way to detect the orientation of a face presented as input to the memory. A face with a strong negative value on the second eigenvector would be categorized as a profile view, and a face with a strong positive value as a frontal view.

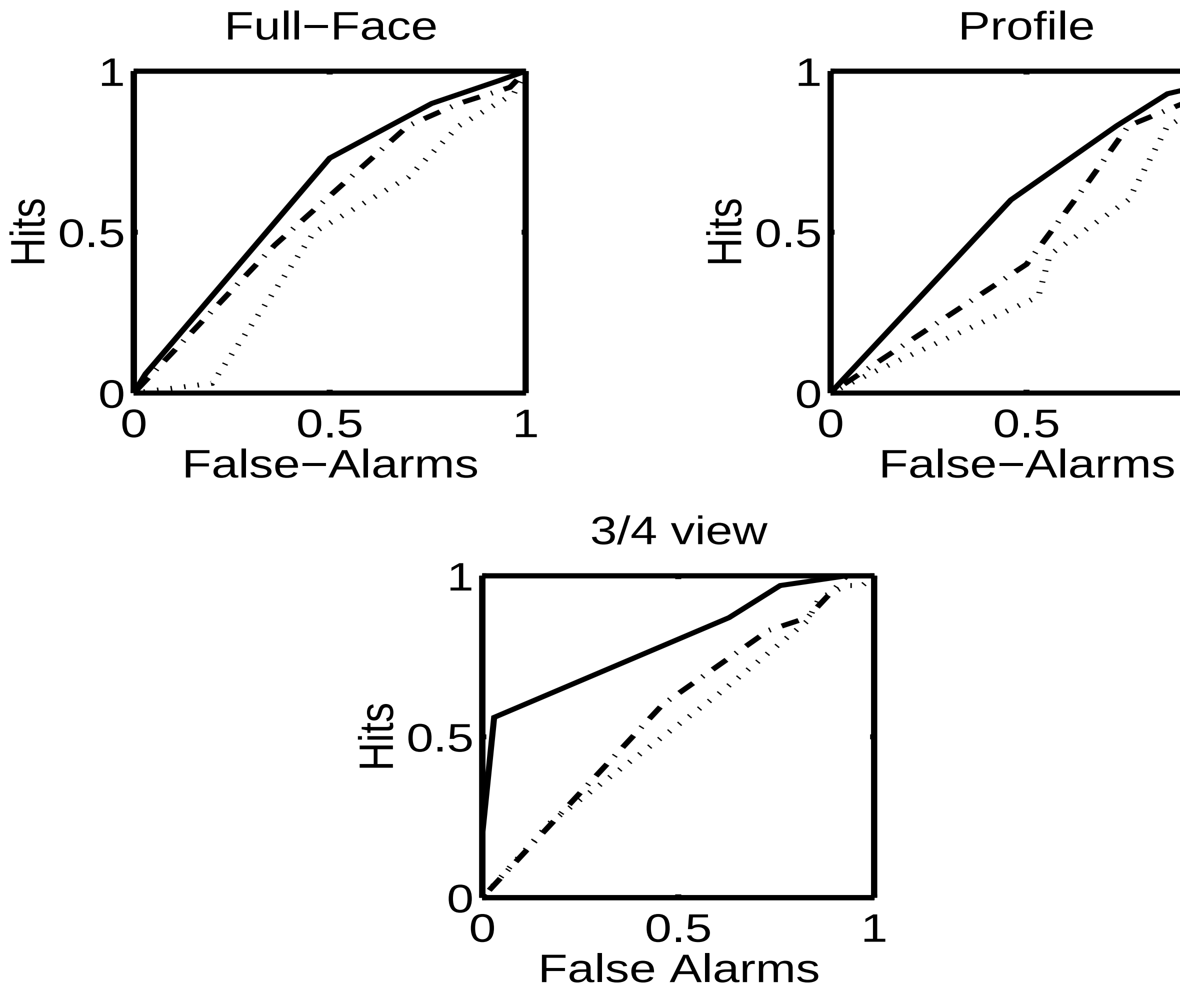

FIGURE 6. ROC curve as a function of learning condition and view angle. The dotted lines represent the 1-view condition, the dashed lines the 4-view condition, and the solid lines the 9-view condition.

Likewise, a face with an intermediary negative value would be categorized as an intermediary orientation between profile and 3/4, and a face with a positive value as an intermediary orientation between frontal and 3/4 view. Finally, a face with a zero value would be likely to be a 3/4 view. As an illustration of the rôle of the second eigenvector for coding the orientation of the faces, Figure 8 shows that, when the first two eigenvectors are combined by addition, a profile view of a face appears. Inversely, if we subtract the first two eigenvectors we create a frontal view.

From a theoretical point of view, the dissociation between orientation-specific and identity-specific eigenvectors can be analyzed in terms of both the eigenvalues associated with the eigenvectors and the visual properties of face images. First, the eigenvalues of the face autoassociative memory correspond to the variance of the projections of the faces onto the eigenvectors. The eigenvector with the largest

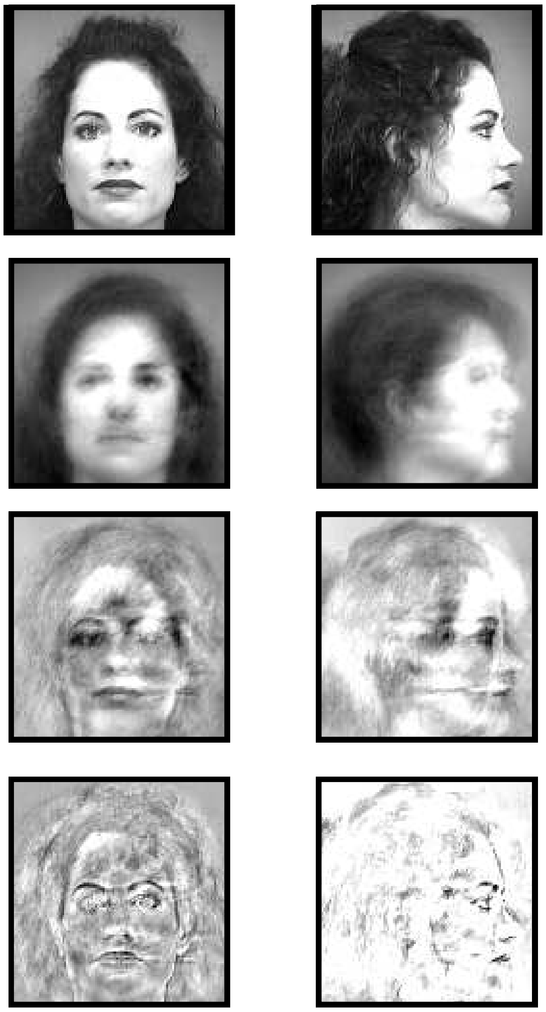

FIGURE 7. A frontal and a profile view of a face reconstructed from top to bottom with 1) all the eigenvectors; 2) the first ten eigenvectors; 3) eigenvectors 11 to 50 and 4) the last 100 eigenvectors of a face autoassociative memory.

eigenvalue accounts for the largest proportion of variance in the face set, the eigenvector with the second largest proportion of variance accounts for the second largest proportion of variance and so on. Therefore, the eigenvectors with large eigenvalues capture the perceptual information that is common to many faces and those with small eigenvalues capture the perceptual information shared by only a few faces. Second, because faces are highly similar objects—they all share the same features arranged in roughly the same configuration—the correlation between two faces in the same orientation (*e.g.,* two frontal views) is larger than the correlation between two views of a given face (*e.g.,* full-face and profile). Therefore, the information relative to the orientation of the faces accounts for more variance in the pixel-by-pixel autoassociative memory than the information relative to the identity of the faces. Consequently, it is captured by the eigenvectors with the largest eigenvalues. Finally, since the eigenvectors are orthogonal, the eigenvectors with smaller eigenvalues capture the residual information, which is the highly detailed information useful to discriminate between individual faces.

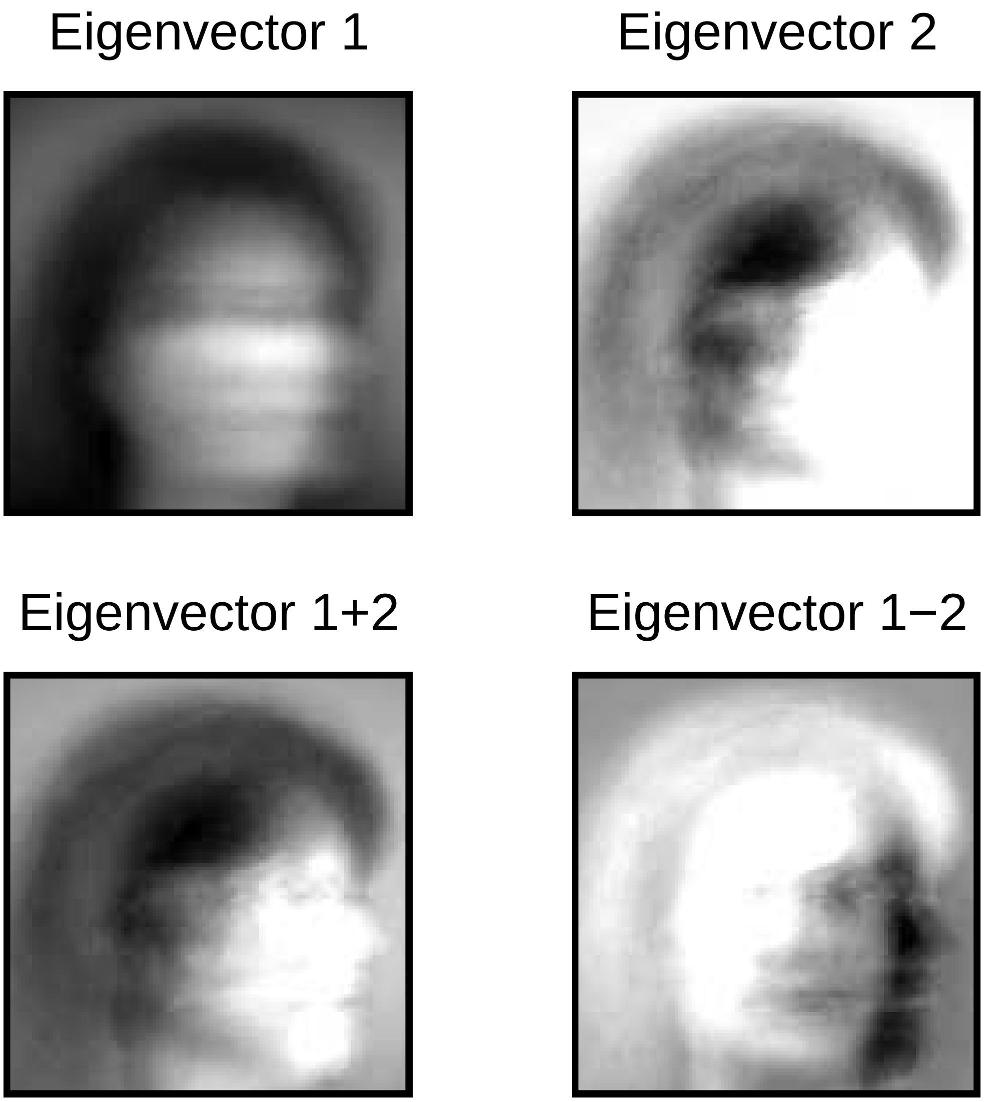

FIGURE 8. Illustration of the *rôle* of the second eigenvector in coding the orientation of the faces. Adding the second eigenvector to the first one gives rise to a profile view, subtracting it gives rise to a frontal view.

4. Conclusion

The study reported here supports the proposition that it is not necessary to use an explicit 3D representation of faces for a simple linear autoassociator to recognize faces from new view angles. Faces can be represented by a limited set of 2D pixel-based images. The only constraint is that the views need to be spaced closely enough for the model to interpolate between two views and transfer to new views.

An interesting byproduct of using 2D pictures of faces in conjunction with an autoassociative memory is a spontaneous dissociation of information relative to the position of the face and the identity of the face. This dissociation is reminiscent of previous dissociations found between semantic components (*e.g.,* gender, race) of face and identity (cf. [20]). The method of displaying eigenvectors and combining them, as shown in this paper, can, in principle, be applied to any eigenvector based methods, and to materials other than faces.

A last point worth noting is that this study is by nature exploratory and so we did not try to optimize the performance of the model. Further studies are needed to examine whether the performance of the model can be improved by using more faces or different sets of multiple views. We can imagine, for example, that using a set of views sampling the rotation of the head with unequal steps (smaller steps near the critical orientations of 0 and 90 degrees and greater steps for intermediate orientations) might help increase the memory's ability to generalize to new views. Also, taking into account the symmetrical aspect of faces might help improving the generalization performance of the model. Finally, it is possible that the problem of recognizing faces from a wide range of orientations is not a linear problem, and therefore a nonlinear algorithm (or classification network operating on an eigenvector representation of the faces) would yield better recognition performance than both the signal detection approach used here or the nearest neighbor algorithm used by Pentland *et al.*[23].

References

[1] H. Abdi, and D. Valentin, "Modèles neuronaux, connexionistes et numériques pour la mémoire des visages", *Psychologie Française*, Vol. 39, pp. 375–391, 1994.

[2] D. Valentin, H. Abdi, A.J. O'Toole, and G.W. Cottrell, "Connectionist models of face processing: A survey", *Pattern Recognition*, Vol. 27, pp. 1209–1230, 1994.

[3] M. Turk, and A. Pentland, "Eigenfaces for recognition", *J. Cognitive Neurosci.*, Vol. 3, pp. 71–86, 1991.

[4] W.D. Dukes, and W. Bevan, "Stimulus variation and repetition in the acquisition of naming responses", *J. Exper. Psychol.*, Vol. 74, pp. 178–181, 1967.

[5] J.C. Bartlett, and J.E. Leslie, "Aging and memory for faces versus single views of faces", *Memory and Cognition*, Vol. 14, pp, 371–381, 1986.

[6] D. Marr, and H.K. Nishihara, " Representation and recognition of the spatial organization of three dimensional shape", *Proc. Royal Soc. of London, B*, Vol. 200, pp. 269–294, 1978.

[7] E. Rosch, "Principle of categorization". In E. Rosch and B. Lloyd (Eds.), *Cognition and categorization*, Hillsdalle Erlbaum, pp. 27–48, 1978.

[8] S. Edelman, and H. Bülthoff, "Orientation dependence in the recognition of familiar and novel views of three dimensional objects", *Vision Research*, Vol. 32, pp. 2385–2400, 1992.

[9] D.G. Lowe, "Three-dimensional object recognition from single two-dimensional images", *Artificial Intelligence*, Vol. 31, pp. 355–395, 1987.

[10] S. Ullman, "Aligning pictorial descriptions: An approach to object recognition", *Cognition*, Vol. 32, pp. 97–136, 1989.

[11] M.J. Tarr, and S. Pinker, "Mental rotation and orientation dependence in shape recognition", *Cognitive Psychol.*, Vol. 21, pp. 233–282, 1990.

[12] S. Edelman, and D. Weinshall, "A self-organizing multiple-view representation of 3-D objects", *Biol. Cybern.*, Vol. 64, pp. 209–219, 1991.

[13] H. Bülthoff, and S. Edelman (1992), "Psychological support for a two dimensional view interpolation theory of object recognition", *Proc. National Academy of Sci.*, Vol. 89, pp. 60–64, 1992.

[14] T. Poggio, and S. Edelman, "A network that learns to recognize three-dimensional objects", *Nature*, Vol. 343, pp. 263–266, 1990.

[15] A. Abdi, "A generalized approach for connectionist autoassociative memories: interpretation, implications and illustration for face processing", In J. Demongeot (Ed.) *Artificial Intelligence and Cognitive Sciences*, Manchester University Press, 1988.

[16] T. Kohonen, *Associative Memory: A System Theoretic Approach*, Springer-Verlag, 1977.

[17] A.J. O'Toole, R.B. Millward, and J.A. Anderson, "A physical system approach to recognition memory for spatially transformed faces", *Neural Network*, Vol. 1, pp. 179–199, 1988.

[18] A.J. O'Toole, H. Abdi, K.A. Deffenbacher, and J.C. Bartlett, "Classifying faces by race and sex using an autoassociative memory trained for recognition", In K.J. Hammomd and D. Gentner (Eds.) *Proc. 13th Annu. Conf. Cognitive Sci. Soc.*, Hillsdale Erlbaum, 1991.

[19] L. Sirovich, and M. Kirby, "Low-dimensional procedure for the characterization of human faces", *J. Opt. Soc. Am. A*, Vol. 4, pp. 519–554, 1987.

[20] A.J. O'Toole, H. Abdi, K.A. Deffenbacher, and D. Valentin, "A low-dimensional representation of faces in higher dimensions of the space", *J. Opt. Soc. Am. A*, Vol. 10, pp. 405–411, 1993.

[21] H. Abdi, D. Valentin, B. Edelman, and A.J. O'Toole, "More about the difference between men and women: Evidence from linear neural networks and the principal component approach", *Perception*, Vol. 24, pp, 539–562, 1995.

[22] H. Murase, and S.K. Nayar, "Learning and recognition of 3D objects from appearance", *IEEE 2nd Qualitative Vision Workshop*, 1993.

[23] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition" *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 1994.

[24] A. Abdi, *Les réseaux de neurones*, Grenoble: Presses Universitaires de Grenoble, 1994.

[25] A. Abdi, "A neural network primer", *J. Biol. Syst.*, Vol. 2, pp. 247–281, 1994.

[26] D. Valentin, H. Abdi, A.J. O'Toole, "Categorization and identification of human face images by neural networks: A review of the linear autoassociative and principal component approaches", *J. Biol. Syst.*, Vol. 2, pp. 413–429, 1994.

[27] D.M. Green, and J.A. Swets, *Signal Detection Theory and Psychophysics*, New York: Wiley, Reprinted by Krieger, Huntington: NY, 1966.

[28] J.A. Anderson, J.W. Silverstein, S.A. Ritz, and R.S. Jones, "Distinctive features, categorical perception, and probability learning: Some applications of a neural model", *Psychol. Rev.*, Vol. 84, pp. 413–451, 1977.