Spanning the face space

Shimon Edelman
Center for Biological and Computational Learning
MIT E25-201
Cambridge MA 02142, US
edelman@ai.mit.edu

Revised, January 16, 1997

Abstract

The paper outlines a computational approach to face representation and recognition, inspired by two major features of biological perceptual systems: graded-profile overlapping receptive fields, and object-specific responses in the higher visual areas. This approach, according to which a face is ultimately represented by its similarities to a number of reference faces, led to the development of a comprehensive theory of object representation in biological vision, and to its subsequent psychophysical exploration and computational modeling.

Keywords: receptive fields, ensemble encoding, similarity, face space, prototypes

1 Introduction

The human perceptual system performs the task of face recognition with remarkable efficiency: we can recognize a large number of individuals, and compensate seemingly effortlessly for variations in viewing direction, illumination conditions, and expressions. In the recognition of objects other than faces, the compensation for these factors presents a challenge known in psychology as the problem of object constancy [16, 17]. For such objects, different views of the same shape may look sufficiently distinct from one another to defeat simplistic approaches to recognition based on storage and recall of raw images [41, 15].

Efficient algorithms that address some of the problems involved in achieving object constancy have been developed in the past decade [41, 39, 21, 42]. These algorithms are typically based on the notion of normalization or alignment: the stored model, or the input image, or both, are transformed so as to maximize the fit between the two, and the model that yields the closest fit is declared to be recognized. The normalization approach, however, may not be the best choice in the recognition of faces, where the problem of constancy is aggravated by the need to distinguish between objects that are very similar to each other to begin with, and which are prone to become less distinguishable following the normalization step. To further complicate the situation, snapshots of different faces seen under similar viewing conditions are likely to resemble each other more closely than images of the same person, taken under different viewing conditions (cf. [23] and Figure 1; note that this is not true of general objects: two arbitrary shapes will probably look as *unlike* each other as any two other shapes, whereas two arbitrary faces tend to look *like* each other — especially to an observer of a different race [3]).

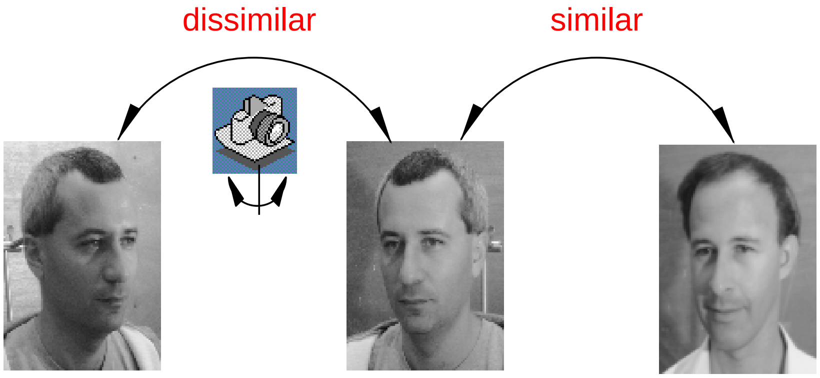


Figure 1: Recognition by direct comparison of face images is not likely to work, because different views of the same person may be less similar to each other than images of different people taken under similar viewing conditions.

These considerations underscore the importance of the development of face representation under which images of the same person would be unconditionally more similar to each other than images of different people [44]. Note that this requirement is, basically, twofold, because it combines constancy and discriminability constraints, which can be satisfied — or violated — independently of each other. In the present paper, I review some of the computational approaches to recognition in the light of those constraints. The structure of the paper follows initially that of a generic face recognition system, which starts with the computation of a representation designed to remain constant under changing viewing conditions (section 2), and proceeds to determine the identity of the face from the resulting representation (section 3). An alternative approach to face recognition emerges, however, as the discussion progresses; this approach aims for achieving constancy and discriminability within a common computational framework, based on representing faces by their similarities to a chosen set of face prototypes (section 4). The development of this approach to face representation can be traced by following references [8, 9, 14, 19].

2 Face representation

2.1 Geometric measurements

A system capable of computing a description of the geometrical properties of a viewed object from its image can obviously form its invariant representation. The route to invariance, or object constancy, via the recovery of the geometry of the object thus has been traditionally appealing to theorists of vision. In face representation, the geometric approach is usually taken to consist of the recovery of two kinds of features, some coding the shapes of face parts such as the eyes and the mouth, others — the locations of these parts within the face [48, 30].

Face recognition algorithms based on geometric features have been shown promising [4]. There are, however, two problems with this approach to face representation. In computer vision, the recovery of the geometric features from raw images appears to be too unreliable to be useful in practice in systems that are required to operate without human intervention. At the same time, it is difficult to envisage a biologically plausible implementation of geometric feature recovery algorithms that could be integrated into a theory of face processing in human vision.

2.2 Receptive fields

A biologically credible alternative to geometric features can be found in the approaches based on convolving the image with a bank of receptive fields (filters); see Figure 2. According to these approaches, which mimic the initial stages of the mammalian visual pathway, the representational substrate consists of numerous filters, with relatively large and overlapping supports (receptive fields), spread over the image. The spatial weighting function of the individual filters is, typically, shallow and graded (i.e., the changes of the weight over the support of the receptive field are gradual rather than abrupt, and are moderate). The image is represented by a vector of numbers, each corresponding to the correlation between the weighting profile of one of the filters and the portion of the image that falls under its support.

A representational system based on a bank of filters can be designed to meet the two objectives mentioned above — constancy and discriminability — each of which is equally important in face processing. First, the profile of the filter can be chosen for maximum invariance of the output with respect to viewing conditions such as illumination (cf. [44]; invariance with respect to viewpoint is typically sought at the subsequent stages of processing, as explained below). Second, the graded profile and the high degree of overlap between adjacent filters can be adjusted to optimize the spatial resolution supported by the system, and, with it, the discriminability of the different faces. As demonstrated recently, these properties of filter-based systems can lead to hyperacuity-level performance, in which spatial detail considerably smaller than the size of the individual receptive fields can be resolved in the output of the system [34, 45]. This property of filter-based systems is clearly of great importance for representing faces (and for discerning among facial expressions), which frequently differ by minute details from one another.

The feasibility of filter-based representation of faces has been demonstrated in the recognition system of [14], which was among the first to employ this approach to face processing (at about the same time, a similar approach has been applied in computer vision for stereo processing [18]; of course, it has been employed also in biological visual systems for a long time [10]). More recently, a system whose first stage follows the principle of filter-based representation has been shown to perform very well in face recognition [47, 46].

2.3 Holistic features

How do the geometric and the filter-based approaches to face representation fare as psychological theories? Only a few of the dozens of studies on the psychology of face recognition address directly issues that can be meaningfully related to computational models of object processing. One such study is the extensive investigation of the psychology of face representation undertaken by Rhodes [31]. This work concentrated on experimental estimation of the relative importance of "first-order" and "second-order" features of faces (the former being simple features such as the interocular distance, and the latter — relationships between feature values, e.g., the ratio of that distance to the nose length). Importantly, an overwhelming majority of the several dozen features involved in the study were of the kind invariant to changes in the viewing conditions. Thus, the implications of the results (summarized below) extend, in general, to these methods for face representation whose main stress is on invariance (this includes geometric features of [4]). As we shall see, these methods do not fare well as models of feature extraction in face processing in human vision.

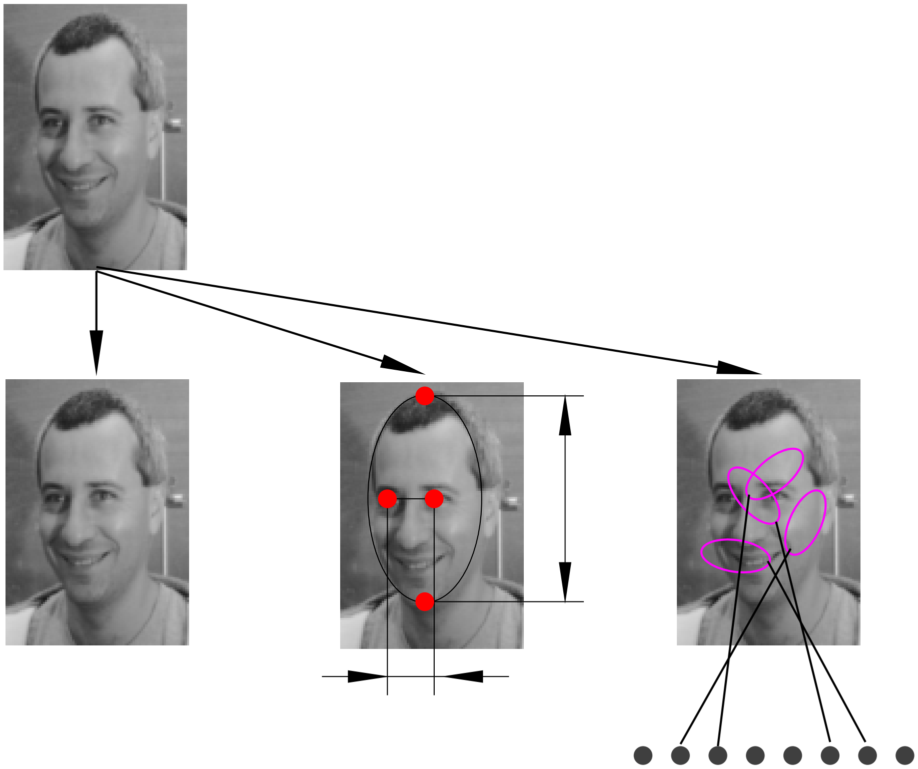


Figure 2: Representing the input face. *Left:* representation by raw images is untenable, because of the effects of viewing conditions, which are likely to cause different images of the same face to appear less similar to each other than images of different faces taken under comparable viewing conditions (see section 1). *Middle:* representation by geometric features (see section 2.1). While such features may provide invariance with respect to viewing conditions, they are rather difficult to extract reliably and automatically from face images. *Right:* representation by activities of receptive fields (see section 2.2). Among the advantages of this approach are the improvement in invariance (compared to raw images) and support for hyperacuity-level resolution.

The study carried out by Rhodes compared the dimensions of perceived similarity between faces (estimated by multidimensional scaling [33] from subjective similarity ratings between pairs of face images) with the various 1st and 2nd-order features of the kind mentioned above. Surprisingly, the winner features were neither 1st-order, nor 2nd-order; rather, two holistic characterizations of the face images, age and sex, emerged as the best predictors of perceived similarity (see [36] for similar results). In other words, the objective measurements that best correlated with the dimensions of variations of face similarity (as perceived by the human subjects and recovered by multidimensional scaling) were neither lengths, nor ratios of lengths, but rather holistic features of faces, whose very extraction from face images is computationally as difficult as recognition itself.

Note that this outcome speaks against the idea of face representation relying on geometric features, but is not incompatible with the approach based on a large number of graded-profile overlapping receptive fields. This is because representation of the latter kind conforms to the principle of least commitment [22], according to which stimulus information is not to be discarded if it may be needed later in the processing stream. By their very nature, geometric features violate this principle (e.g., the presence of an edge element in a certain location of the image is an all-or-none feature which constitutes a commitment *par excellence*), whereas a vector of activities of receptive fields, by virtue of the smooth input-output function it implements, is likely to preserve and pass on as much information as possible [11]. As we shall see in section 4, a second processing stage built on top of a filter-based representation will provide a direct account of the results of Rhodes, suggesting a novel approach to settling down the question of the nature of face representation in human vision.

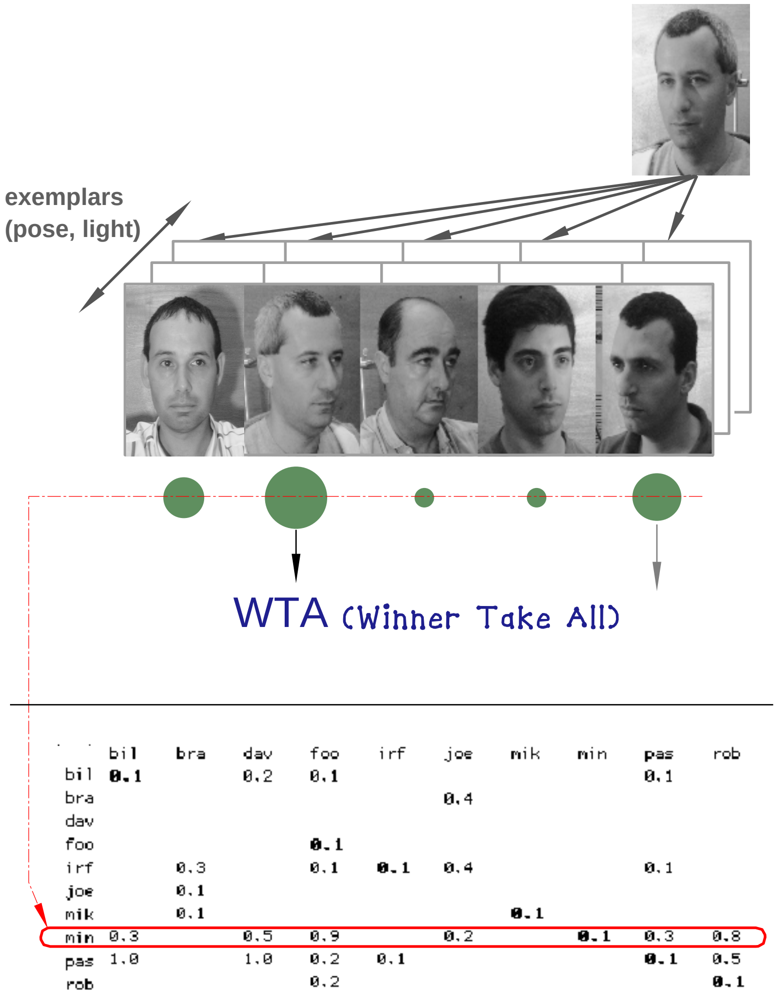
3 Face recognition

3.1 A generic approach

The next step in a face recognition system, following the formation of representations that are as invariant and as informative as possible, is the derivation of the identity of each stimulus face from its representation. Although this step can be implemented in many different ways, I chose to describe here one particular approach: learning the mapping from an intermediate representation to the identity of the face using a neural network classifier. This approach is equivalent to all the other approaches functionally, and it seems to be the most appropriate approach to the modeling of face processing in biological vision.

The structure of a typical system for learning face recognition is illustrated in Figure 3, top. Such a system consists of a number of modules, each containing a neural network classifier trained to recognize images of a particular face. A dedicated module is required for each face to be recognized by the system (note that this is true also for systems in which there is no explicit separation into modules, but which, nevertheless, must be taught or constructed to respond properly to each new face). I shall now describe a particular system of this kind [14], which will serve as a representative example.

In that system, each recognition module was implemented by a radial basis function (RBF)



	bil	bra	dav	foo	irf	joe	mik	min	pas	rob
bil	**0.1**		0.2	0.1					0.1	
bra						0.4				
dav										
foo				**0.1**						
irf		0.3		0.1	**0.1**	0.4			0.1	
joe		0.1								
mik		0.1					**0.1**			
min	0.3		0.5	0.9		0.2		**0.1**	0.3	0.8
pas	1.0		1.0	0.2	0.1				**0.1**	0.5
rob				0.2						**0.1**

Figure 3: *Top:* A generic face recognition system (see section 3.1), composed of a number of modules, each trained to recognize images of a particular face. The outcome of the recognition process in such a system is determined by a Winner Take All operation on the outputs of the individual modules (the size of the disk beneath each module symbolizes the strength of its response). *Bottom:* A confusion table representation of the performance of the generic system (only a part of the entire table is shown, for clarity). Entries along the diagonal correspond to "miss" error rates; off-diagonal entries signify "false-alarm" error rates (zeros omitted for clarity). For instance, the module trained on face bil produced a false alarm on 20% of the views of dav, 10% of the views of foo, and 10% of the views of pas. Note that these confusion values are related to the similarity between bil and the three other faces. As pointed out in section 3.2, discarding the information contained in the off-diagonal entries of a confusion table amounts to giving up valuable clues to the identity of the stimulus.

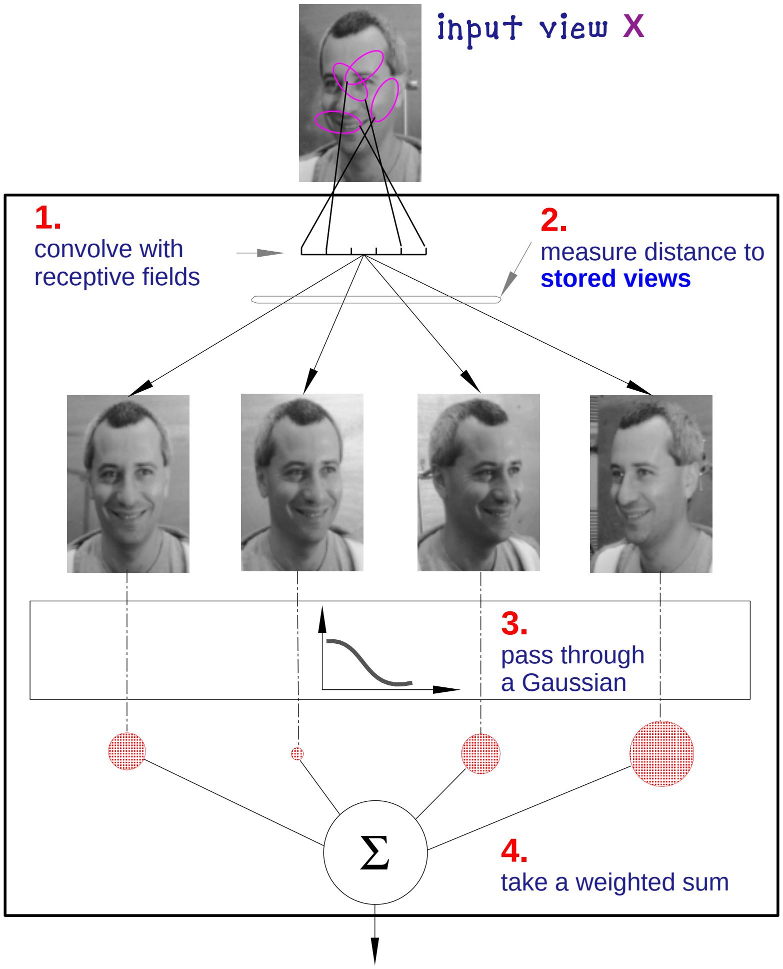


Figure 4: A radial basis function (RBF) module [29, 28], used in [14] to implement the individual recognizer. The input to the module is provided by a bank of filters (see section 2.2). Each of a small number of training views, $\mathbf{v}_t$, serves as the center of a Gaussian basis function $\mathcal{G}(\mathbf{a};\mathbf{b}) = exp\left(\|\mathbf{a}-\mathbf{b}\|^2/\sigma^2\right)$; the response of the module to an input vector $\mathbf{x}$ is computed as $y = \sum_t w_t \mathcal{G}(\mathbf{x};\mathbf{v}_t)$. The weights w_t are learned by computing the least-squares approximate solution of an overdetermined system of linear equations [29]; the spread parameter σ can be learned too, but is usually set to the mean separation of the training vectors. Analogies between this model of recognition and a series of findings in visual neurophysiology (in particular, the reports of view-specific and object-specific cells, both for faces [27] and for other objects [20]) are readily apparent.

network [29, 28], trained to recognize images of one face. Only positive examples were used, that is, the network's parameters were adjusted (by a standard RBF training procedure; see [29]) so that its output (a single real number) was as close as possible to 1 for all the given images of the target face. Images of other faces were not included in the training set, although certain information regarding the entire ensemble of faces with which the system was familiarized was incorporated in the individual modules, via the choice of the RBF spread parameter σ (see Figure 4). This parameter was adjusted to optimize the total performance of all the modules (that is, their acceptance of the target images and their rejection of the non-target ones). The other parameters that were adjusted during training were the weights of the output linear layer of the RBF networks; the centers of the basis functions were placed at the points in the intermediate (filter) representation space corresponding to the training images of the target face.

The recognition system was tested on a subset of the MIT Media Lab database of face images, made available by Turk and Pentland [40], which contained 27 face images of each of 16 different persons. The images were taken under varying illumination and camera location. Of the 27 images available for each person, 17 randomly chosen ones served for training the RBF recognition module, and the remaining 10 were used for testing.

The performance of the individual recognizers was assessed by computing a 16×16 confusion table, in which the entries along the diagonal signified miss rates and the off-diagonal entries — false alarm rates (see Figure 3, bottom). The table was computed row by row, as follows. First, the module for the person whose name appears at the head of the row was trained. Second, the recognition threshold was set to the mean output of the recognizer over the training set, less two standard deviations. Third, the performance of the module on the test images of the same person was computed and the miss rate entered on the diagonal of the table. Finally, the false alarm rates for each recognizer on the 10 test images of the other 15 persons were computed and entered under the appropriate columns of the table. The row by row average (miss and false-alarm) error rate was about 10%, with individually adjusted thresholds.

Note that because of the row by row computation, this figure conveys no information as to the expected performance of a system composed of a number of recognizers (although threshold-setting is useful in the generation of a sparse confusion table). The performance of a complete system would depend on that of an arbitration mechanism capable of deciding in favor of one of the several modules activated (i.e., made to pass the threshold) by the stimulus. In an overwhelming majority of recognition systems, the arbitration is carried out according to the winner-take-all (WTA) principle: an error is declared if the identity of the most active recognizer module is different from that of the stimulus face. In the present case, the total error rate of the WTA step was about 22%, a figure that clearly needs to be improved. As we shall see immediately, a considerable improvement can be achieved simply by postponing the WTA step until more of the information in the distribution of activities of the individual modules is put to use.

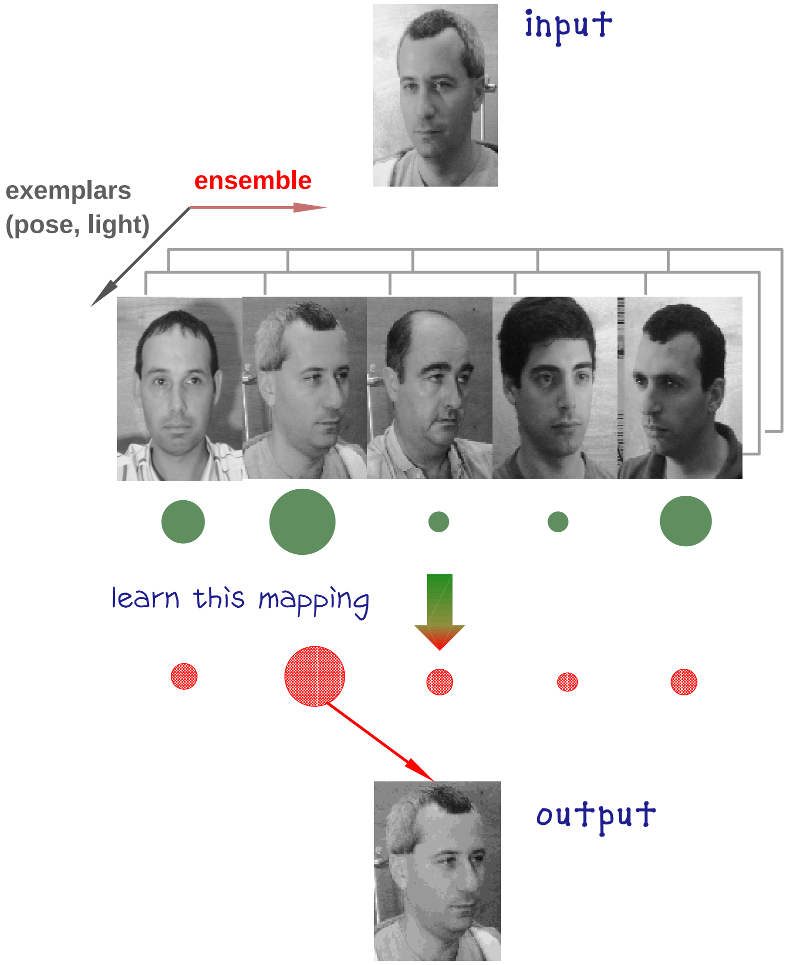


Figure 5: A two-stage approach to face recognition that takes into account ensemble information; see section 3.2.

3.2 Taking advantage of ensemble information

An examination of the confusion table, a part of which is shown in Figure 3, bottom, reveals that many of the individuals tended to be confused with almost any other person in the database. Observe that the pattern of confusion actually carries much information concerning the identity of each stimulus: it is more useful to know the degree to which a face resembles each of an ensemble of other faces, than to single out merely the closest match from that set. To take advantage of this ensemble phenomenon, another RBF module was trained to accept vectors of individual module activities, and to produce vectors of the same length in which the value corresponding to the activity of the correct module was 1, and all other values were 0. The training set for the second-stage RBF module was obtained by pooling the training sets of all 16 first-stage recognizers. The outcome of the recognition of a test image was determined by finding the coordinate in the output vector whose value was the closest to 1. The performance of the two-stage scheme (see Figure 5) was considerably better than that of the individual recognizer stage alone (9% error rate, compared to 22%), demonstrating the importance of ensemble knowledge for recognition. The next section offers an interpretation of this result in view of the psychological evidence in favor of holistic representation of faces [31], and links it to some recent developments in the theory of representation in vision.

4 Discussion

4.1 Face space

The action of the second stage in the recognition system of [14] on a stimulus face amounts to the construction of its representation in terms of similarities to a number of "reference" or prototype faces, which, in turn, are coded holistically (cf. [31]; these are the faces on which the first-stage RBF modules are trained). This representation is much richer than the one from which the winner-take-all decision made by the single-stage version of the system was derived. Specifically, in addition to the identity of the familiar face most similar to the stimulus image, this representation includes also the information regarding the identity of the second, third, etc. most similar face (cf. [6]). The ordering of the familiar faces by their similarity to a novel one can be stored and used to recognize the latter in future encounters (see Figure 7); note that the ranking information should suffice here (just as in nonmetric multidimensional scaling [33, 32] the *ranks* of interpoint distances suffice to determine their locations in the embedding space).

Intuitively, the ensemble encoding of the stimulus image in terms of its similarities (that is, proximities, or inverse distances) to a set of reference faces may be likened to representing a point on a terrain in terms of its distances to a set of landmarks, as it is done when triangulation is used in navigation or in cartography. This analogy leads naturally to the notion of *face space*, which has been brought up in a number of discussions of face processing in the past [40, 43, 2], and which acquires a special meaning in the context of the present scheme for face representation.

The particular conception of face space proposed here is based on a more general approach to visual representation, delineated in [12]. Consider the image of a face presented to a recognition system as a point in a multidimensional *image space*, where each dimension corresponds to the value of one of the receptive fields ("pixels").[1] When the face rotates in front of the eye, or when the illumination shifts, the point ascribes a generically smooth trajectory in the image space (the trajectory is smooth if the viewing transformation is, and if the function describing the action of the receptive field on the image is differentiable). Note that two different faces will give rise to two trajectories running generally in parallel in the image space, to the extent that the two faces are similar (more on this in section 4.2).

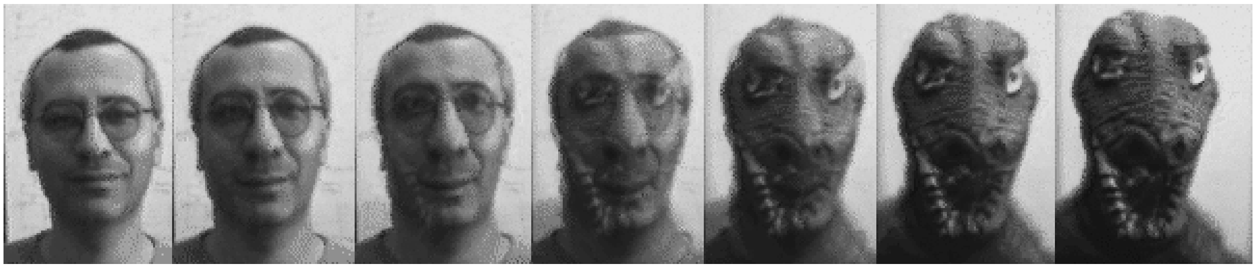


Figure 6: Morphing corresponds to a smooth movement in the image space between points corresponding to different faces (see section 4.1).

Now, a classifier module of the kind illustrated in Figure 4 effectively interpolates the image-space trajectory corresponding to the different views of the same face, from a number of views given to it in training (in an RBF implementation, these views serve as the centers of the basis functions). As a result, the output of an individual module fed with a new image can be interpreted

[1] This is only possible if the images are brought into register [1], at least roughly. In the system of [14], the registration was accomplished by an affine transformation, guided by correspondence between eye and mouth locations in the images to be matched. These locations were identified automatically, using a symmetry-seeking operator [30].

as the proximity between that image and the entire trajectory [12].

Note that morphing between images of two similarly oriented faces (Figure 6) corresponds to a smooth trajectory through the image space. The interpolation of *this* trajectory from examples is, however, more complicated than interpolation of the view space of an individual face, for the simple reason that the required output here is not a scalar (in particular, it does not make sense merely to output a constant value for the different possible inputs in this case). For the purpose of recognition, it suffices to represent a morphing trajectory in the vector space defined by the outputs of a number of modules acting in parallel. Each point on the trajectory is then the representation of the input image in terms of its proximities (similarities) to the set of stored reference faces, as described in the preceding section. The second stage of a recognition system can then act on this representation, not by interpolation, but rather by associating with it a desired identity label (see Figure 7).

Another way to look at the action of a number of modules each of which is coarsely tuned to a particular face is by considering the modules as implementing receptive fields in the face space. Recall that the two figures of merit of a representation, mentioned in the introduction, were constancy and discriminability; under the above interpretation, each individual module contributes constancy with respect to viewpoint to the emerging representation (this is what the modules are trained for), while the entire ensemble of modules supports hyper-resolution in the face space (which is analogous to the spatial hyperacuity afforded by a system of graded overlapping receptive fields in low-level vision).

4.2 Class-based processing

The relatively large number of views needed to train an RBF module appears to be a disadvantage of the proposed scheme for face representation, when compared to the ability of human subjects to generalize recognition to a novel view of a face previously seen from a limited range of viewpoints. A remedy to this problem has been proposed in the form of class-based processing — a concept intended to account for the limited ability of humans to generalize recognition to novel views of inverted faces, compared to nearly ideal generalization for upright faces [24]. Assuming that the visual system stores information regarding the appearance of a considerable number of (upright) faces under a relatively wide range of conditions, it should be able to use such information to

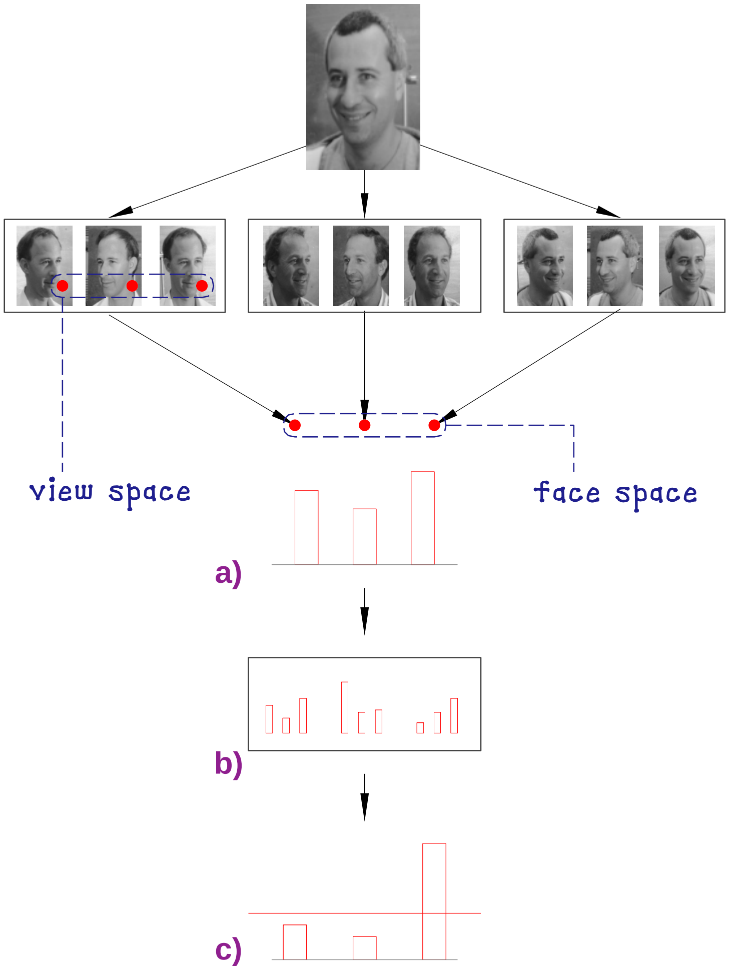


Figure 7: how to use the two-stage system in recognition. *Step* **a***:* represent the stimulus in terms of its similarities to a number of reference faces (with the view-induced variation neutralized prior to the computation of the similarity, e.g., using a view interpolation mechanism such as an RBF module; see Figure 4). *Step* **b***:* compare the similarity-based representation of the stimulus to a number of such representations stored in memory. *Step* **c***:* decide which of the memory traces fits the stimulus the best. It has been conjectured [8, 9] that a similar sequence of processes is involved in shape representation in primate vision, where findings of cortical columns tuned to a variety of objects have been reported [37, 38].

generalize to novel views of unfamiliar faces. In terms of the notion of a face space discussed above, the trajectories corresponding to objects that belong to the same class would lie sufficiently close together to allow interpolation by a specially trained subsystem. Thus, given experience with a certain class of objects (say, upright faces), the system should be able to generalize well from single views of members of that class, but not of another class (say, inverted faces; note that in the image space points corresponding to upright and inverted images of the same face lie far apart). A computational scheme taking advantage of this observation has been described in [19]; recently, it has been applied to the modeling of face recognition by human subjects [25, 26], with promising results.

4.3 Summary

I have described a computational approach to face representation and recognition which takes as its starting point a system implemented in 1991 [14]. Although the performance of that system is not impressive by the present-day standards, its approach proved to be fruitful in a number of respects. In particular, it led to the development of a comprehensive theory of object representation in biological vision [8, 9], and to its subsequent psychophysical exploration and computational modeling [7, 5, 12, 35, 13]. Additional issues suggested by this approach are currently under exploration. Some of these issues are the extension of the theory to a variety of object classes as well as a number of levels of processing, and the study of its implications for biological vision by electrophysiological and functional brain imaging methods.

Acknowledgments

Parts of this work were carried out while SE was at the Weizmann Institute of Science, Rehovot, Israel. Some of the empirical results mentioned in this paper were obtained in cooperation with D. Reisfeld and Y. Yeshurun. I am grateful to S. Ullman for prompting me to think about the computational significance of those results back in 1992, and to T. Poggio for pointing out to me the importance of solving the correspondence problem as a prerequisite for treating images as a vector space. Betty Edelman, Hervé Abdi, and two anonymous reviewers provided constructive comments on the manuscript. Thanks also to Alice O'Toole, for rekindling my interest in human face processing, even as it seemed to me that the computational mystery behind this topic was largely gone.

References

[1] Beymer, D., and Poggio, T., Image representations for visual learning. *Science* **272** (1996) pp. 1905–1909.

[2] Bichsel, M., and Pentland, A., Human face recognition and the face image set's topology. *Computer Vision, Graphics, and Image Processing: Image Understanding* **59** (1994) pp. 254–261.

[3] Brigham, J. C., The influence of race on face recognition. In *Aspects of face processing*, ed. by Ellis, H. D., Jeeves, M. A., and Newcombe, F., (Martinus Nijhoff, Dordrecht, 1986) pp. 170–177.

[4] Brunelli, R., and Poggio, T., Face recognition through geometrical features. In *Proc. 2nd European Conf. on Computer Vision, Lecture Notes in Computer Science*, ed. by Sandini, G., **588** (Springer Verlag, 1992) pp. 792–800.

[5] Cutzu, F., and S. Edelman, S., Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Science* **93** (1996) pp. 12046–12050.

[6] Duvdevani-Bar, S., and Edelman, S., On similarity to prototypes in 3D object representation. CS-TR 95-11, Weizmann Institute of Science, 1995.

[7] Edelman, S., Representation of similarity in 3D object discrimination. *Neural Computation* **7** (1995) pp. 407–422.

[8] Edelman, S., Representation, Similarity, and the Chorus of Prototypes. *Minds and Machines* **5** (1995) 45–68.

[9] Edelman, S., Representation is representation of similarity. CS-TR 96-08, Weizmann Institute of Science, 1996. submitted to Behavior and Brain Sciences.

[10] Edelman, S., Receptive fields for vision: from hyperacuity to object recognition. In *Vision*, ed. by Watt, R. (MIT Press, Cambridge, MA, 1997), in press.

[11] Edelman, S., Vision reanimated. In *Proc. 7th Rosenön Workshop on Computer Vision*, ed. by Aloimonos, Y., Carlsson, S., and Eklundh, J.-O., (L. Erlbaum, Hillsdale, NJ, 1997), forthcoming.

[12] Edelman, S., Cutzu, F., and Duvdevani-Bar, S., Similarity to reference shapes as a basis for shape representation. In *Proceedings of 18th Annual Conf. of the Cognitive Science Society*, ed. by Cottrell, G., pp. 260–265, San Diego, CA, July 1996.

[13] Edelman, S., and Duvdevani-Bar, S., Similarity, connectionism, and the problem of representation in vision. *Neural Computation* **9** (1997) pp. 701–720.

[14] Edelman, S., Reisfeld, D., and Yeshurun, Y., Learning to recognize faces from examples. In *Proc. 2nd European Conf. on Computer Vision, Lecture Notes in Computer Science*, ed. by Sandini, G. **588** (Springer Verlag, 1992) pp. 787–791.

[15] Edelman, S., and Weinshall, D., Computational approaches to shape constancy. In *Perceptual constancies: why things look as they do*, ed. by Walsh, V., and Kulikowski, J., (Cambridge University Press, Cambridge, UK, 1997), in press.

[16] Ellis, R., Allport, D. A, Humphreys, G. W, and Collis, J., Varieties of object constancy. *Q. Journal Exp. Psychol.* **41A** (1989) pp. 775–796.

[17] Jolicoeur, P., and Humphrey, G. K., Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In *Perceptual constancies*, ed. by Walsh, V., and Kulikowski, J., (Cambridge University Press, Cambridge, UK, 1997), in press.

[18] Jones, D. G., and Malik, J., A computational framework for determining stereo correspondence from a set of linear spatial filters. In *Proc. 2nd European Conf. on Computer Vision, Lecture Notes in Computer Science*, ed. by Sandini, G. **588** (Springer Verlag, 1992) pp. 395–410.

[19] Lando, M., and Edelman, S., Receptive field spaces and class-based generalization from a single view in face recognition. *Network* **6** (1995) pp. 551–576.

[20] Logothetis, N. K., Pauls, J., and Poggio, T., Shape recognition in the inferior temporal cortex of monkeys. *Current Biology* **5** (1995) pp. 552–563.

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

[22] Marr, D., Early processing of visual information. *Phil. Trans. R. Soc. Lond. B* **275** (1976) pp. 483–524.

[23] Moses, Y., Adini, Y., , and Ullman, S., Face recognition: the problem of compensating for illumination changes. In *Proc. ECCV-94*, ed. by Eklundh, J.-O. (Springer Verlag, 1994) pp. 286–296.

[24] Moses, Y., Ullman, S., and Edelman, S., Generalization to novel images in upright and inverted faces. *Perception* **25** (1996) pp. 443–462.

[25] O'Toole, A. J., and Edelman, S., Face distinctiveness in recognition across viewpoint: An analysis of the statistical structure of face spaces. In *Proc. 2nd Intl. Conf. on Face and Gesture Recognition*, ed. by Essa, I. (1996) pp. 10–15.

[26] O'Toole, A. J., Edelman, S., and Bülthoff, H. H., Face recognition and identification from novel viewpoints. MPIK TR 31, Max Planck Institut für biologische Kybernetik, Tübingen, Germany, June 1996.

[27] Perrett, D. I., Mistlin, A. J., and Chitty, A. J., Visual neurones responsive to faces. *Trends in Neurosciences* **10** (1989) pp. 358–364.

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

[29] Poggio, T., and Girosi, F., Regularization algorithms for learning that are equivalent to multilayer networks. *Science* **247** (1990) pp. 978–982.

[30] Reisfeld, D., Wolfson, H., and Yeshurun, Y., Detection of interest points using symmetry. In *Proceedings of the 3rd International Conference on Computer Vision* (IEEE, Washington, DC, 1990) pp. 62–65.

[31] Rhodes, G., Looking at faces: first-order and second-order features as determinants of facial appearance. *Perception* **17** (1988) pp. 43–63.

[32] Shepard, R. N., Metric structures in ordinal data. *J. Math. Psychology* **3** (1966) pp. 287–315.

[33] Shepard, R. N., Multidimensional scaling, tree-fitting, and clustering. *Science* **210** (1980) pp. 390–397.

[34] Snippe, H. P., and Koenderink, J. J., Discrimination thresholds for channel-coded systems. *Biological Cybernetics* **66** (1992) pp. 543–551.

[35] Sugihara, T., Edelman, S., and Tanaka, K., Representation of objective similarity among 3D shapes in the monkey. *Invest. Ophthalm. Vis. Sci. Suppl. (Proc. ARVO)* **37** (1996) abstract.

[36] Tanaka, J. W., and Farah, M. J., Parts and wholes in face recognition. *Quarterly J. Exp. Psychol.* **46A** (1993) pp. 225–245.

[37] Tanaka, K., Inferotemporal cortex and higher visual functions. *Current Opinion in Neurobiology* **2** (1992) pp. 502–505.

[38] Tanaka, K., Neuronal mechanisms of object recognition. *Science* **262** (1993) pp. 685–688.

[39] Thompson, D. W., and Mundy, J. L., Three-dimensional model matching from an unconstrained viewpoint. In *Proceedings of IEEE Conference on Robotics and Automation* (1987) pp. 208–220, Raleigh, NC.

[40] Turk, M., and Pentland, A., Eigenfaces for recognition. *J. of Cognitive Neuroscience* **3** (1991) pp. 71–86.

[41] Ullman, S., Aligning pictorial descriptions: an approach to object recognition. *Cognition* **32** (1989) pp. 193–254.

[42] Ullman, S., and Basri, R., Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13** (1991) pp. 992–1005.

[43] Valentine, T., Representation and process in face recognition. In *Vision and visual dysfunction*, ed. by Watt, R., (Macmillan, London, 1991) **14** pp. 107–124.

[44] Weiss, Y., and Edelman, S., Representation of similarity as a goal of early visual processing. *Network* **6** (1995) pp. 19–41.

[45] Weiss, Y., Edelman, S., and Fahle, M., Models of perceptual learning in vernier hyperacuity. *Neural Computation* **5** (1993) pp. 695–718.

[46] Wiscott, L., Fellous, J.-M.,, Krüger, N., and von der Malsburg, C., Face recognition and gender determination. In *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, IWAFGR'95* (Zurich, 1995) pp. 92–97.

[47] Würtz, R. P., Vorbrüggen, J. C., and von der Malsburg, C., A transputer system for the recognition of human faces by labeled graph matching. In *Parallel Processing in Neural Systems and Computers*, ed. by Eckmiller, R., Hartmann, G., and Hauske, G. (North Holland, Amsterdam, 1990) pp. 37–41.

[48] Yuille, A. L.,, Hallinan, P. W., and Cohen, D. S., Feature extraction from faces using deformable templates. *International Journal of Computer Vision* **3** (1992) pp. 99–112.