Receptive field spaces and class-based generalization from a single view in face recognition

Maria Lando Shimon Edelman[†*]
Department of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel
† edelman@wisdom.weizmann.ac.il

August 29, 1995

Abstract

We describe a computational model of face recognition, which generalizes from single views of faces by taking advantage of prior experience with other faces, seen under a wider range of viewing conditions. The model represents face images by vectors of activities of graded overlapping receptive fields (RFs). It relies on high spatial frequency information to estimate the viewing conditions, which are then used to normalize (via a transformation specific for faces), and identify, the low spatial frequency representation of the input. The class-specific transformation approach allows the model to replicate a series of psychophysical findings on face recognition, and constitutes an advance over current face recognition methods, which are incapable of generalization from a single example.

*To whom all correspondence should be addressed.

1 Introduction

The ability to recognize a novel view of a face previously seen under a restricted range of conditions is one of the more amazing feats of human vision. We describe a computational model of generalization from a single view in face recognition, built around the assumption that such generalization is made possible by the previous experience of the visual system with similar objects (i.e., other faces). In particular, we assume that the visual system stores information regarding the appearance of a considerable number of faces under a relatively wide range of conditions, and ask how such information can be put to use in generalizing to novel views of an unfamiliar face. The answer suggested by our results provides an explanation of a number of recent psychophysical findings, and may lead to a significant enhancement in the performance of the present-day computer vision systems for face recognition, which are by and large incapable of generalization from a single view.

1.1 Psychophysical background

A new insight into the computational basis of human generalization performance in face recognition has been achieved as a result of a recent study that compared generalization for upright and inverted faces (Moses et al., 1993). In that study, human subjects performing an upright face discrimination task were found to generalize nearly perfectly to face images obtained under novel illumination and viewpoint direction. In comparison, for inverted faces, the generalization to novel views was significantly worse, even though the experimental conditions required that the subjects discriminate between familiar views of the same inverted faces with high reliability, before they were allowed to proceed to the testing stage.

One may observe that images of inverted faces are statistically identical to those of upright faces on the pixel level. Poor generalization performance on inverted faces thus constituted an important control, suggesting that the generalization mechanism in face recognition is object-specific, rather than universal (i.e., valid for all images). Furthermore, the subjects' ability to generalize from a single image of an upright face made the possibility of strictly model-based generalization (which, in principle, requires more that one image to form a full 3D model) unlikely. Moses et al. concluded that the generalization occurs at an intermediate or class-based level, where upright faces constitute a class distinct from that of inverted faces (despite the pixel by pixel equivalence in complexity of the two types of stimuli).

The extensive experience of the human visual system with upright faces in everyday life constitutes the central difference between upright and inverted faces (Diamond and Carey, 1986). Presumably, it is this experience that allows upright, but not inverted, faces to be recognized easily under a wide range of unfamiliar conditions (see Figure 1). Indeed, proficiency in face recognition

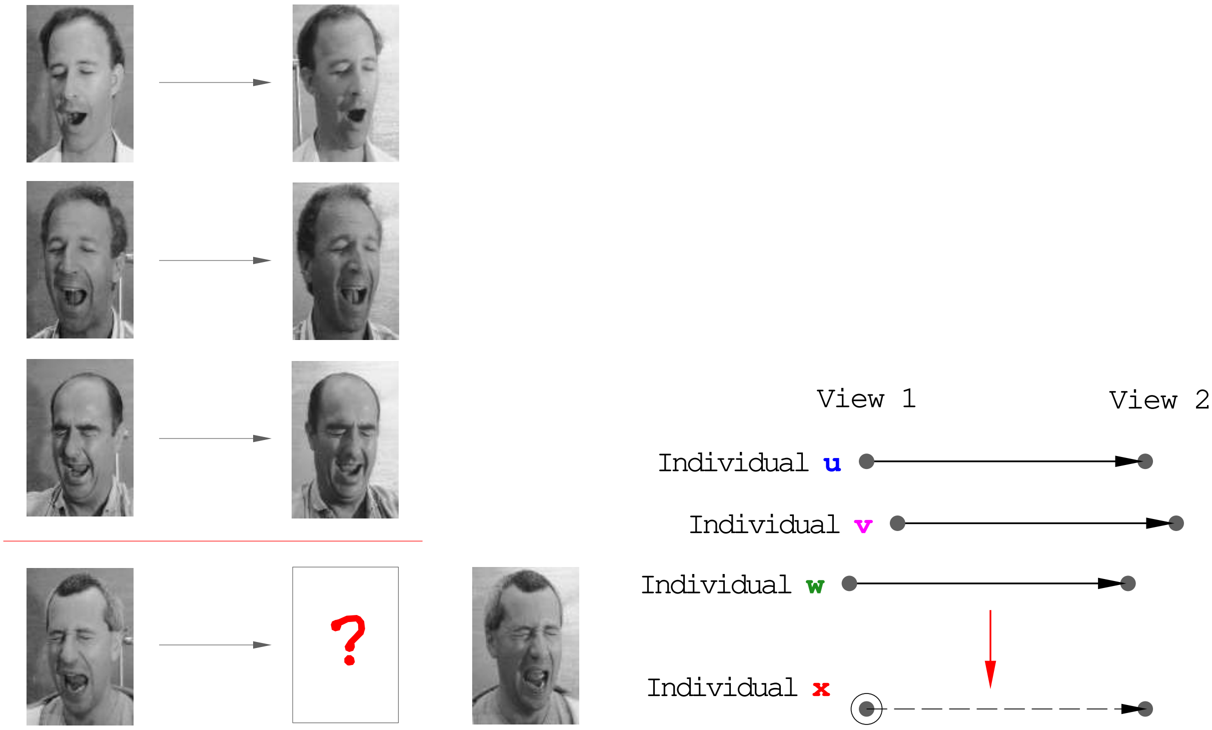


Figure 1: An illustration of the idea of class-based processing. *Left:* Experience with a number of shapes belonging to the same class (in this case, the class of faces) undergoing a certain transformation can serve as a basis for the generalization of that transformation to a new member of the same class of shapes (i.e., a new face). *Right:* Whereas in computer graphics applications the goal of this operation is to generate the *image* of the new face under the specified transformation (Beymer et al., 1993), recognizing a face from an unfamiliar viewpoint calls merely for the normalization of its representation in some feature space that preserves face identity (e.g., the space spanned by properly chosen receptive fields); see section 1.2.1.

appears to be, to a considerable extent, an acquired ability. Children's recognition of upright faces improves steadily from age 6 to 10, dips temporarily between ages 11 and 12, then climbs to an adult level (Carey et al., 1980). In comparison, recognition performance on inverted faces does not change throughout the life and remains significantly worse (Carey and Diamond, 1977; Carey et al., 1980). Moreover, subjects who, following training, successfully *recognize* inverted faces under fixed viewing conditions still perform relatively poorly when required to *generalize* to a novel viewpoint or even a novel illumination (Moses et al., 1993).

1.2 Computational background

1.2.1 Related work

The main purpose of the present work is to formalize the intuitive notions of perceptual experience and of class-based generalization, outlined in the preceding section. At least two different computational approaches to these issues has been suggested recently. The first of these (Basri, 1992) concentrates on the relationship between classification and recognition, and assumes the availability of a library of 3D models of prototypical objects. In the first stage of the recognition process in Basri's system, the prototypes are aligned (Ullman, 1989) with the input image. Alignment here is class-based in the sense that the transformation between the best-matching prototype and the input is taken to apply to the entire class of shapes which the prototype represents. In the second stage, this transformation is reused to align the individual members of that class with the image.

The second approach, due to Poggio and Vetter (1992), avoids the need for a library of 3D models. In their work, Poggio and Vetter show how class-specific transformations of 3D objects can be learned from examples of 2D object views. For objects consisting of "clouds" of points in 3D and represented by 2D views obtained by projection, Poggio and Vetter define the notion of a linear class (i.e., linear combinations of a basis set of objects). Because of linearity, a transformed (e.g., rotated) version of an object that belongs to a linear class is a weighted sum of similarly transformed basis objects with the same coefficients, and the same relationship holds for object views. Poggio and Vetter mention a possible extension of this approach from projections of points to images of surfaces using texture mapping. In a subsequent work, Beymer et al. (1993) determine the transformation that relates two images of a face using an optic flow algorithm, and apply this transformation to generate a similarly transformed image of novel face, from a single available view.

1.2.2 The present approach

Objects are said to belong to the same class if they are more similar to each other than to other objects. Thus, we consider similarity to be of primary importance to the present work, and require that a biologically credible definition of similarity between face images be provided before proceeding to model class-based processing in face recognition in human vision. The biological constraint on similarity is not entirely compatible with the coordinate-based features and the pixel-based image representations mentioned above, which do not fit well the computational characterization of mechanisms of biological information processing. Consequently, we have adopted the following guidelines in developing our approach to class-based processing.

Representation by receptive fields. In biological visual systems, a natural basis for the definition of similarity can be derived from the concept of processing units with localized receptive fields (RFs). The RFs of the primary visual cortex correspond to the psychophysically defined spatial frequency channels of (Wilson and Bergen, 1979); activities of the graded-profile highly overlapping RFs at the previous stages is the only input available to any processing stage in the visual pathway past the retina. The overlap between the constituent RFs has been shown to improve the utility of a representation (Snippe and Koenderink, 1992). This improvement, however, saturates when the number of RFs reaches a few hundreds, making a relatively small set of RFs nearly as useful as a full dimensionality-preserving coverage of the retinal space (Weiss and Edelman, 1995).

Clustering by face identity. We assume that the metrics of the internal representation space for faces reflects the true metrics prevailing in the objective "face space," in the sense that different appearances of the same face (i.e., views taken under different viewing positions and illumination conditions) tend to cluster together. This assumption relies on a recent computational investigation, which found that images of the same face form tighter clusters at the higher levels of an RF-based representation hierarchy resembling the one found in mammalian vision (Weiss and Edelman, 1995).

1.2.3 Class-based generalization

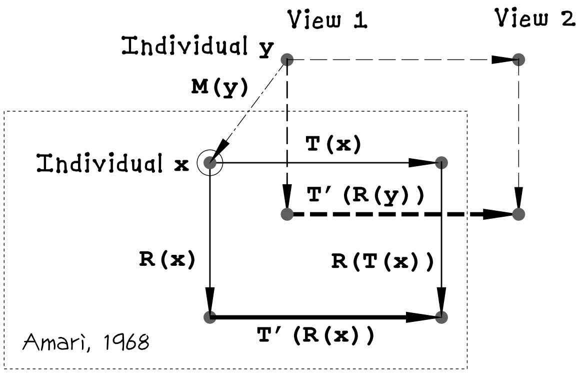


Figure 2: Effective feature-space representation of objects that may undergo certain transformations requires that those transformations commute with the feature extraction process (solid arrows; see Amari, 1968). In this adaptation of Amari's commutation diagram, the similar effect of transformation on the representations of different faces is taken to signify actual RF-space parallelism of the two transformation vectors, shown by thick lines. Legend: `M` is morphing between different faces; `R` is transduction by a bank of receptive fields; `T` is transformation between different views.

We assume that applying the same transformation (change of viewpoint or illumination) to images of different faces results in similar changes in the internal (RF-space) representation of each face. Thus, the transformation that maps a prototypical view of a face to its appearance under specific viewing conditions is assumed to be similar across different faces. This assumption can be illustrated by a simple diagram shown in Figure 2, in which the similarity of the normalizing transformations across faces is expressed geometrically, by the parallelism of the corresponding vectors in the representation space. Note the relationship between this diagram and Amari's (1968, 1978) illustrations of the desirable properties of feature-space representations. As pointed out by Amari, effective feature-space representation of objects that may undergo certain transformations requires that those transformations, in a sense, commute with the feature extraction process. In the simplest case, the commutation property is satisfied merely because, in the appropriate representation space, transformation between fixed views corresponds to translation in a fixed direction, irrespective of the object identity.

The assumption of similarity of the normalizing transformations across faces constitutes the main working hypothesis of the present paper.[1] Consider two viewing positions, V_0 and V_1. We conjecture that a face previously seen from V_0 can be recognized when seen from V_1 because the system stores a snapshot of the RF activity evoked by the exposure to the face in viewing position V_0, and compares the stored activity vector to the present one, taking into account experience with numerous other faces in both positions, V_0 and V_1. Note that the same approach can be applied to the case of varying illumination, face expressions and different combinations of these parameters.

2 Class-specific transformations in RF space

To substantiate the notion of class-based generalization, we examine the expected behavior of RF-space representations of face images under various transformations of faces, and compare the resulting computational predictions with data derived from artificial (computer graphics) and real (human) faces.

2.1 A quantification of geometrical similarity between faces

We start by defining geometric similarity of human faces, via a shape variation parameter δ that expresses the upper bound across individuals on the changes in each component of the normal at any point of a face. We assume that the normals $\mathbf{N}_{f_1}$ and $\mathbf{N}_{f_2}$ at two corresponding points of two faces f_1 and f_2 satisfy the inequality

[1] The real "commutation diagram" turns out, however, to be more complicated than the simple illustration discussed above; see Figure 12.

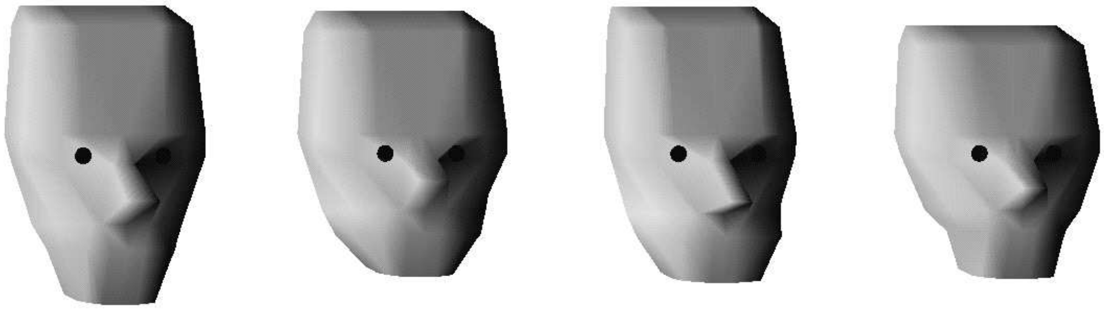


Figure 3: Faces obtained by random variation of the parameters in the face geometry model (see section 2.3).

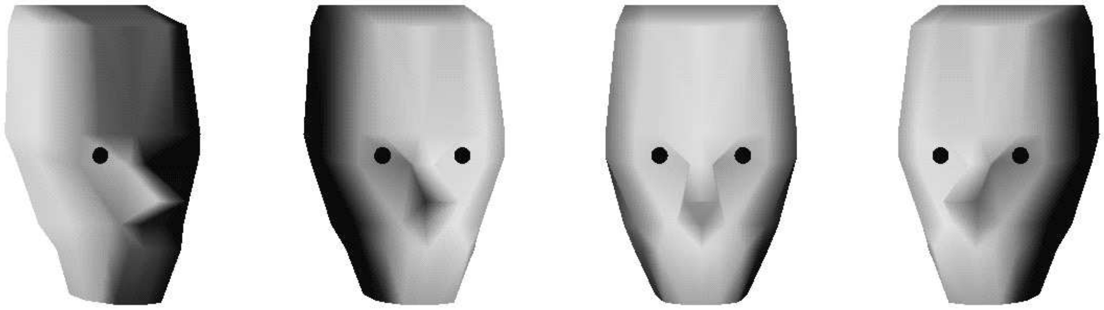


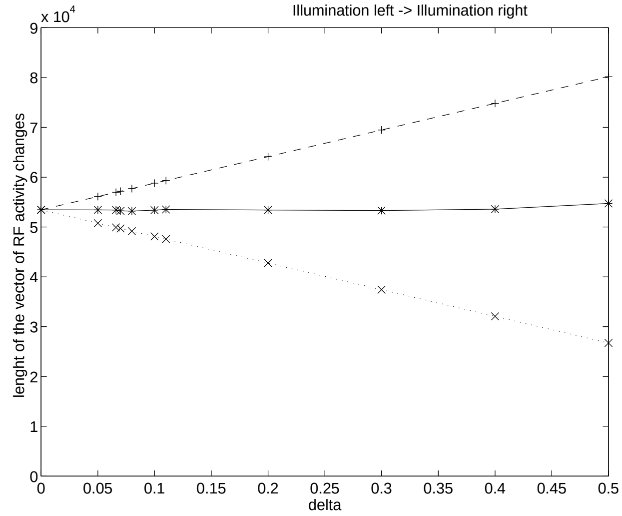
Figure 4: One of the face models rendered under different viewpoint and illumination conditions.

$$N^i_{f_1}(1-\delta) \;\leq\; N^i_{f_2} \;\leq\; N^i_{f_1}(1+\delta)$$

in each of their components N^i. Thus, the resemblance of different faces to a certain average one can be described in terms of a single parameter δ. We further assume that this geometric similarity is properly reflected in the representational space of the visual system, namely, in the similarity of the vectors of RF activity changes between given viewing conditions for different faces.

2.2 Theoretical predictions

The similarity assumption can be substantiated, by using the shape parameter δ to place a bound on the RF-space transformation of face representations, caused by changes in the viewing conditions (a detailed derivation is given in the appendix). The activity of each of the (linear) units that span the RF space is proportional to the intensity around the projection of some patch of the surface of the viewed object, integrated over the unit's receptive field. Assuming that the reflectance function of the surface is predominantly diffuse (as it is the case for faces), the activity of each RF will depend



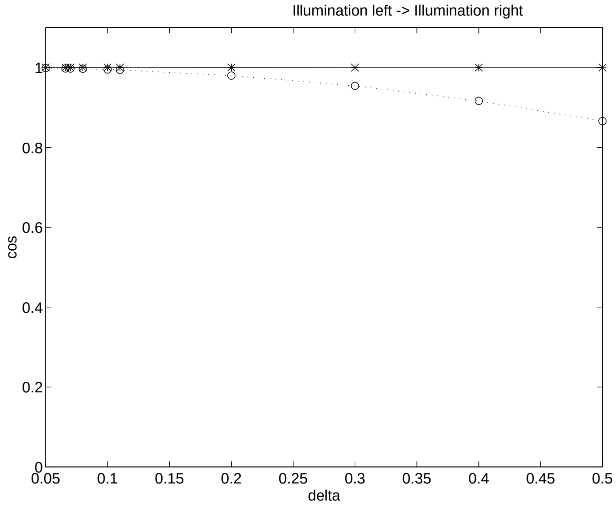


Figure 5: The behavior of the RF activity changes, caused by shifting the illumination source from the left to the right side of the face. plotted against the similarity parameter δ of the synthetic faces. *Left:* lengths of the vectors for different values of δ, and the predicted lower and upper bounds; *Right:* the angles between the vectors of RF activity changes, and the predicted lower bound.

on the direction of the normal at the corresponding point on the surface, relative to the illumination vector. The RF activity pattern will thus change if either the orientation of the face (and with it the directions of the normals), or the illumination change. We now address these two cases separately.

Consider first the effect of a change in illumination conditions. Let $\mathbf{X}^{(f_1)} \in R^k$ and $\mathbf{X}^{(f_1)} + \Delta\mathbf{X}^{(f_1)}$ be the RF-space representations of a face f_1 under two different illuminations. The i'th component of $\Delta\mathbf{X}^{(f_1)}$ is then proportional to the change in the normal direction at the i'th RF, and can differ at most by δ from person to person. As shown in the appendix, for two persons f_1 and f_2 for whom the corresponding normals differ at most by δ, the length of the difference vector $\|\Delta\mathbf{X}^{(f_1)} - \Delta\mathbf{X}^{(f_2)}\|$ is bounded from above by δ. Furthermore, the cosine of the angle between $\Delta\mathbf{X}^{(f_1)}$ and $\Delta\mathbf{X}^{(f_2)}$ is bounded from below by $\sqrt{1-\delta^2}$.

Consider now the effect of a shift in the viewpoint. The vector $\Delta\mathbf{X}^{(f_1)} - \Delta\mathbf{X}^{(f_2)}$ defined in the previous paragraph depends also on the interaction between changes in the illumination direction and in the viewpoint. The length of the difference vector and the lower bound on the cosine of the angle between $\Delta\mathbf{X}^{(f_1)}$ and $\Delta\mathbf{X}^{(f_2)}$ thus depend both on δ and on the minimal change of the projection of the normal onto the illumination direction, caused by a shift in the viewing position (i.e., $\min \|(\mathbf{L} \cdot \Delta\mathbf{N})\|$, where $\mathbf{L}$ is the illumination vector and $\mathbf{N}$ is the normal vector). In the appendix, we show that the lengths of the vectors of RF activity changes can differ at most by $2\delta / \min|(L \cdot \Delta N)|$, and that the bound on the cosine of the RF-space angle between $\Delta\mathbf{X}^{(f_1)}$ and

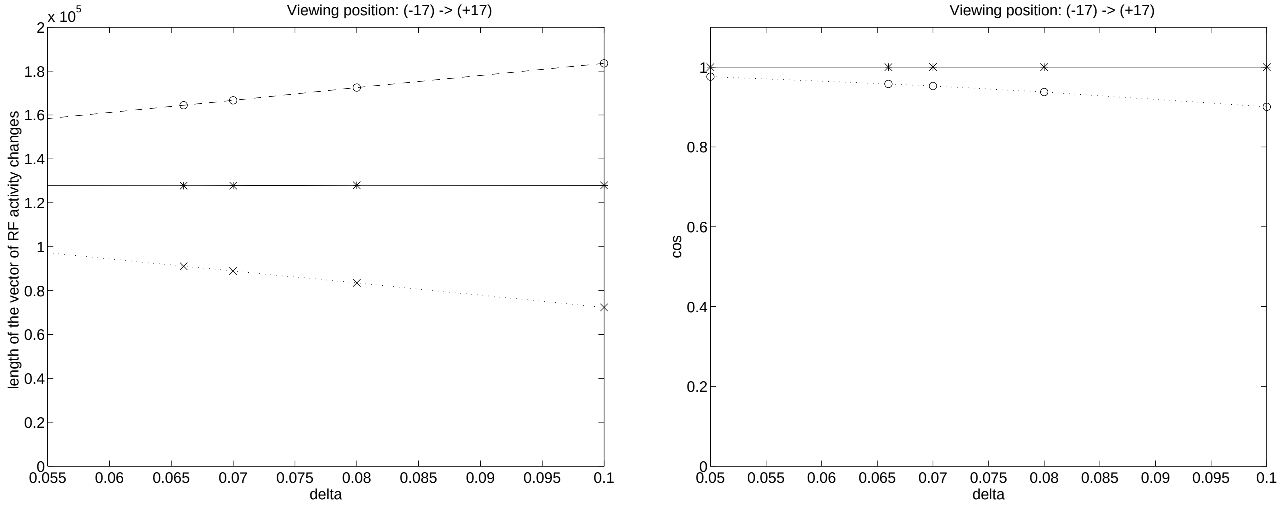


Figure 6: The dependency of the length of the vector of RF activity change (left) and the cosine of the angle between the vectors of RF activity changes (right) on the similarity parameter δ. The minimal change in the projection of the normal on illumination direction equals 0.46.

$\Delta\mathbf{X}^{(f_2)}$ is $\sqrt{1-(2\delta/\min|(\mathbf{L}\cdot\Delta\mathbf{N})|)^2}$.

2.3 Class-specific transformations for artificial faces

We have subjected the bounds on class-specific transformations in RF space, derived in the previous section, to empirical testing. To obtain a family of face shapes with controlled values of the geometric similarity parameter δ, we constructed a stylized model of a generic human face using computer graphics. The model, controlled by 19 experimentally determined geometrical parameters, was built using a 3D graphics toolkit (SGI Inventor). Individual differences in face shape were represented by randomly varying the parameters, within limits imposed by the value of δ. A range of values of δ around several percent allowed the system considerable flexibility in representing faces with different shapes. Figure 3 depicts four of the 50 faces that were obtained by varying the 19 parameters. Each of the 50 face models was rendered under five different viewing positions and three illumination directions, and the resulting images were represented by activities of 500 RFs (Weiss and Edelman, 1995).

We found that the values of $\|\Delta\mathbf{X}^{(f_i)}-\Delta\mathbf{X}^{(f_j)}\|$ and $\angle(\Delta\mathbf{X}^{(f_i)},\Delta\mathbf{X}^{(f_j)})$ calculated using the synthetic faces were nearly identical for the different face models (i.e., for different choices of f_i, f_j), and in all cases were well within the analytically derived bounds (see Figures 5 through 7). Note that in the case of a shift in the viewing position, the theoretically predicted bounds depend not only on δ but also on the minimal change in the projection of the normal onto the illumination direction. We

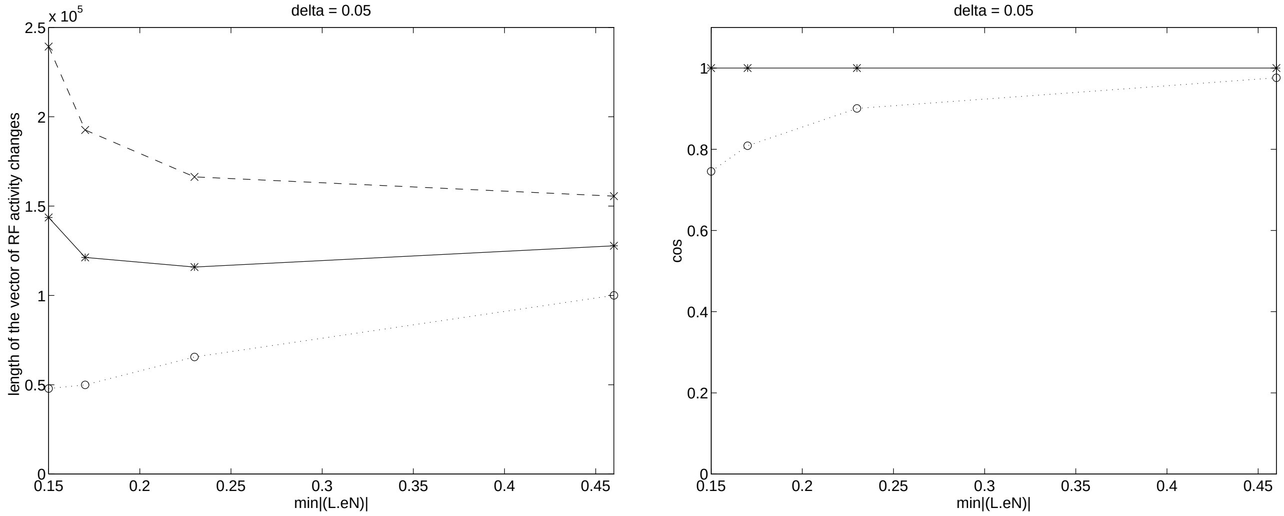


Figure 7: The dependency of the length of the vector of RF activity changes (left) and the cosine of the angle between the vectors of RF activities (right) on the minimal change in the projection of the normal on the illumination direction. The similarity parameter of the faces is equal to 0.05.

found that for our model this value depends only on the particular type of viewing condition change, and not on the identity of the face. When the viewpoint shifted from -17^o to $+17^o$ with respect to the frontal view, the minimal[2] change in the projection of the normal onto the illumination direction was found to be equal to 0.46; the resulting dependence of $\|\Delta\mathbf{X}^{(f_i)} - \Delta\mathbf{X}^{(f_j)}\|$ and $\angle(\Delta\mathbf{X}^{(f_i)}, \Delta\mathbf{X}^{(f_j)})$ on δ is shown in Figure 6. Another way to examine this dependence is by looking at the effect of shifting the viewpoint by different amounts, for a fixed value of $\delta = 0.05$ (see Figure 7).

2.4 Class-specific transformations for real faces

We next assessed the behavior of $\|\Delta\mathbf{X}^{(f_i)} - \Delta\mathbf{X}^{(f_j)}\|$ and $\angle(\Delta\mathbf{X}^{(f_i)}, \Delta\mathbf{X}^{(f_j)})$ on real face images, taken from the Weizmann Face Base (Moses et al., 1993).[3] The results for the length and the

[2]The minimum was over all 166 surface patches composing each face model and was the same for all faces.

[3]This database contains twenty images of each of 18 different male faces, without distinctive features (e.g. no glasses, beard, mustache, etc.). All images were taken by the same camera under tightly controlled illumination and viewpoint. The frontal view of all faces were normalized by fixing the location of the face symmetry axis, the external corners of the eyes, and the bottom of the nose, before taking the pictures. A computer-controlled robot positioned the camera at $-34°$, $-17°$, $0°$, $17°$ and $34°$ with respect to the frontal view, in the horizontal plane. The distance of the face from the camera was fixed at about $110cm$. Four distinct illumination conditions were created by turning on and off three fixed light sources. The subjects were asked to assume a neutral expression and to remain still. To reduce the influence of the background, the faces were clipped by an elliptical mask that occluded most of the hair and the neck areas. Each image consisted of 512×352 pixels, 8 bits per pixel.

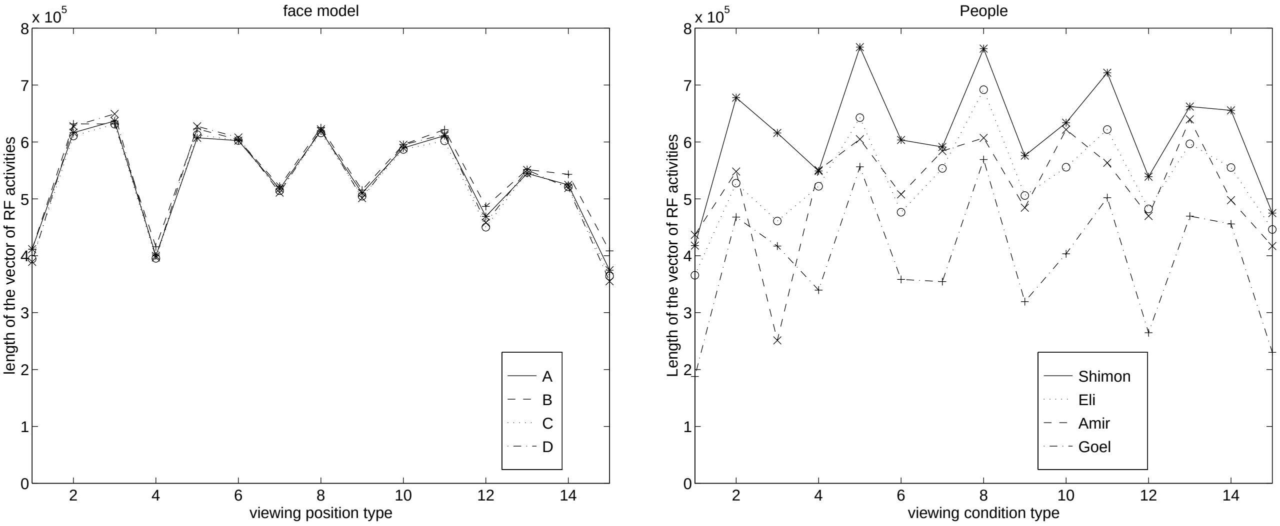


Figure 8: Length changes of the RF-space vectors corresponding to 4 different face models (left) and 4 real faces (right), over 15 different viewing conditions. Note that the changes for different faces are correlated with each other.

angle differences appear, respectively, in Figures 8 and 9, which illustrate the correlated changes induced in the RF-space representations of different faces by changes in the viewing conditions. These results (especially, the tight correlation exhibited by $\angle(\Delta\mathbf{X}^{(f_i)}, \Delta\mathbf{X}^{(f_j)})$ for the different face pairs f_i, f_j) support the hypothesis that view transformations induce similar changes in RF-space representation, regardless of face identity.

3 Class vs. transformation-based clustering in RF space

The results of the previous section provide the foundations for a class-specific approach to face processing, by demonstrating that same-class objects (faces) induce similar patterns in the RF representation space across changes in viewing conditions. This finding can only be put to use in face identification if it is possible to distinguish, in the RF space, between points representing images of the same face from points representing all other faces. In principle, the clustering in the RF space can take one of two possible forms (see Figure 10). The first possibility is that points representing the same face are clustered together. Alternatively, points representing the same viewing conditions can be clustered together. In both cases, the correlations apparent from Figures 8 and 9 would stem from the common displacement of all clusters following a certain transformation (either of view or of face identity).

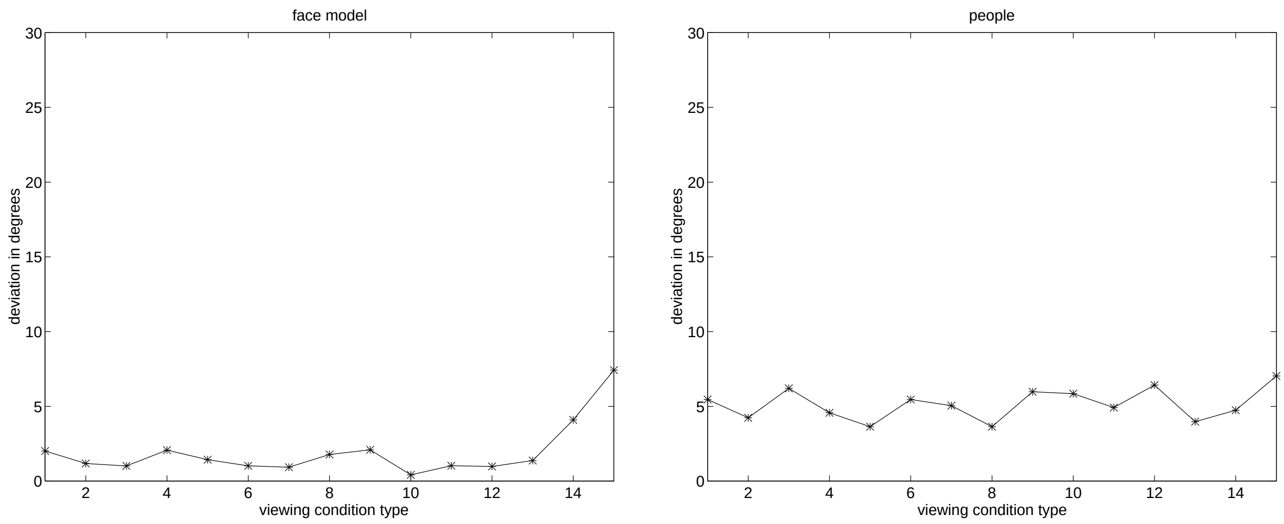


Figure 9: Direction (angle) changes of the RF-space vectors corresponding to 4 different face models (left) and 4 real faces (right), over 15 different viewing conditions. Note the small absolute value of angle changes across viewing conditions, both for synthetic and for human faces.

The spatial frequency analysis of face-difference images reported in (Weiss and Edelman, 1995) suggests that the nature of clustering of faces in RF-space should depend on the size of the RFs used in the representation. We tested this conjecture, by convolving face images with Difference of Gaussian RFs, corresponding in size to the different spatial frequency channels described in (Wilson and Bergen, 1979). Eight values of RF size, between 0.8 cycles per degree (*cpd*) and 16 *cpd*, were used with images of different faces, under different viewing conditions. For each RF size, we conducted 60 trials, with faces chosen randomly from the Weizmann database. The degree of clustering was defined as the ratio R of mean within-cluster and between-clusters distances (under this measure, $R << 1$ indicates good clustering).

Two experiments were carried out. In the first experiment only the illumination was varied (Figure 11, left). Clustering by face identity improved with increasing RF size, while clustering by viewing condition became better when the RF size decreased. In the second experiment, viewing position varied (over 68^o of visual angle) in addition to illumination (Figure 11, right). Here, clustering by face identity did not improve with increasing RF size as much as in the previous case, although it did eventually become better than clustering by viewing position. In particular, the results for the 16 *cpd* filter may indicate that a combination of low and high frequency filters may be used to obtain better clustering performance.[4] This table also provides a comparison of different filters

[4]One way to improve clustering here is by having the system learn from examples an optimal linear combination of low and high frequency filters.

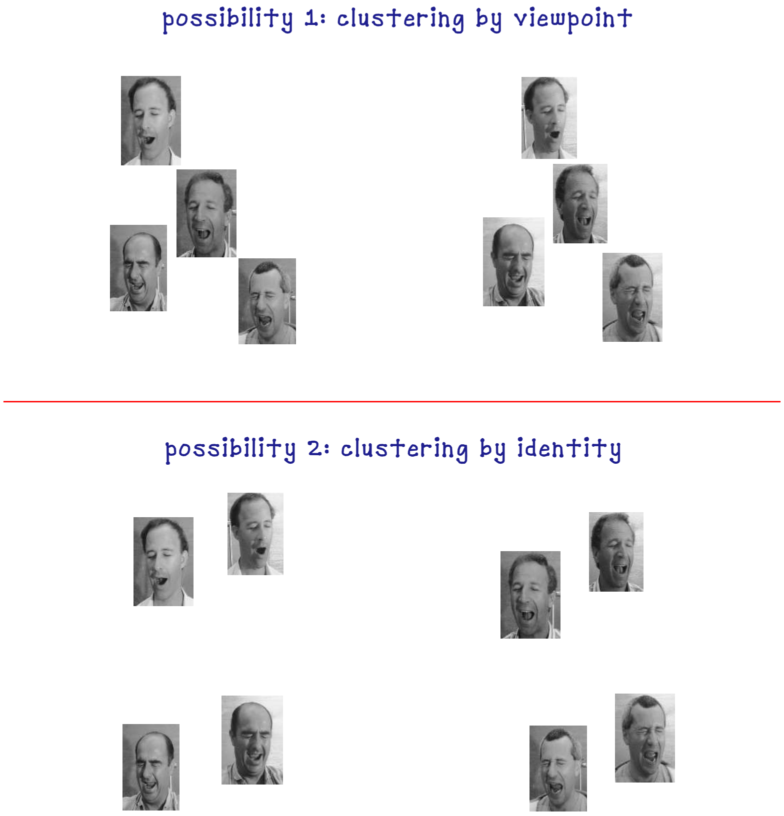


Figure 10: The correlated RF-space changes precipitated by viewpoint and illumination transformations for different faces (namely, the tight bounds on $\|\Delta\mathbf{X}^{(f_i)} - \Delta\mathbf{X}^{(f_j)}\|$ and $\angle(\Delta\mathbf{X}^{(f_i)}, \Delta\mathbf{X}^{(f_j)})$ for different i, j; see section 2) are compatible with two possible manners of clustering of face representations in the RF space. One of these – clustering by face identity – would render the representation more useful for recognition than the other. The actual manner of clustering in spaces spanned by RFs of different sizes is explored in section 3 (see also Figure 11).

and pixel-level representations. Interestingly, the 500-dimensional space spanned by the 16 *cpd* RFs carries nearly the same information about viewing position as the original $512 \times 352 = 180224$ pixel image, demonstrating the redundancy of the pixel-level representation.

The advantage of large (low-frequency) over small RFs in representing face identity may be explained intuitively as follows. Consider first changes in illumination. In our test set, the light source moved from one side of the face to the other, causing a relatively large-scale fluctuation in the gradient of image intensity; this was better averaged out by the larger RFs. Second, when the viewing position changed (causing the face image effectively to slip out from under the RFs positioned over it), the larger RFs had a better chance of continuing to cover more or less the same portion of a face

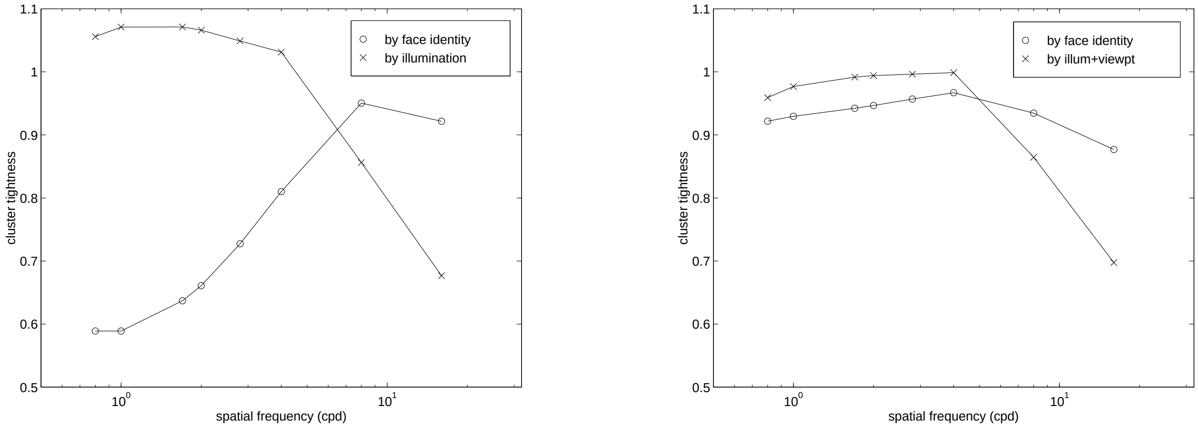


Figure 11: *Left:* The degree of clustering (R; see section 3) by face identity and by illumination, for different RF sizes. *Right:* Clustering by face identity and by viewing conditions (illumination and viewpoint), for different RF sizes.

than the smaller ones.

4 A complete model for face recognition

The principle of class-based generalization, stated in the introduction, requires tighter clustering by identity than that available even with the low spatial frequency RFs. It turns out, however, that one can effectively combine the viewpoint information carried by the high-frequency (16 *cpd*) RFs with the face identity information in the low-frequency (0.8 *cpd*) RFs. In the system described in this section, viewpoint and identity information is extracted from the image, and then combined, by appropriately trained function approximation modules (we used for this purpose radial basis function, or RBF, classifiers (Moody and Darken, 1989; Poggio and Girosi, 1990)). The model consists of the following stages (see Figure 13):

1. **Detection of viewing conditions.** This stage is implemented by an RBF classifier, which accepts image representations in the high-frequency RF space, ${}^{H}\mathbf{X}$, and estimates the viewing conditions $\mathbf{V}({}^{H}\mathbf{X})$.

2. **Normalization to the prototype.** According to the detected viewing conditions $\mathbf{V}({}^{H}\mathbf{X})$., a class-specific transformation transforms the low-frequency RF-space representation ${}^{L}\mathbf{X}$ into a prototypical form ${}^{L}\hat{\mathbf{X}}$ predicted for the input image.

3. **Face identification.** This is implemented by an RBF classifier, which, in the low-frequency RF space, compares the hypothesized prototypical representation ${}^{L}\hat{\mathbf{X}}$ with those of known

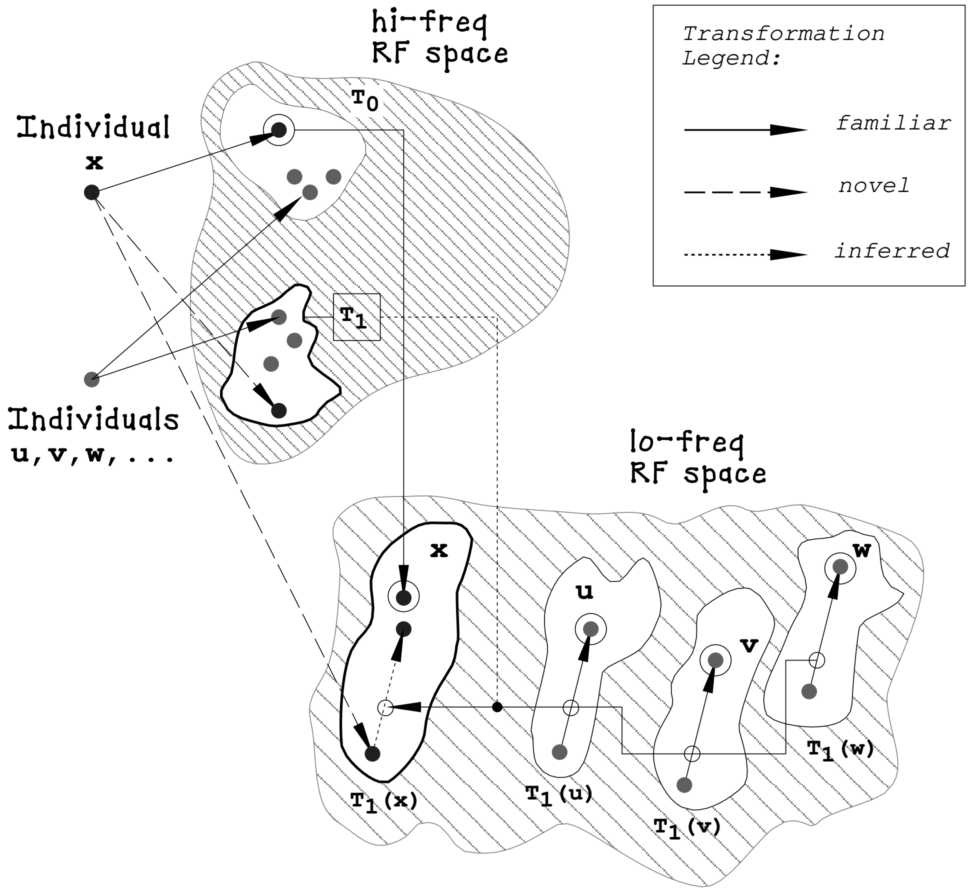


Figure 12: A conceptual scheme for class-based generalization from a single view, described in section 4 (compare with Figure 2). Familiarity with a number of faces (u,v,w,...) under two different viewing conditions (clusters designated by T_0, T_1) helps recognize face x, previously seen only at T_0, from a new view T_1. In the high-frequency RF space, the clustering by view enables the extraction of view information, which is then used to normalize the representation of the input image in the low-frequency RF space. In that space, there may be no clear-cut clustering either by view or by identity, but the normalizing transformations for similar objects (i.e., faces u,v,w,...,x) are similar, which makes the normalization of the novel view of x possible.

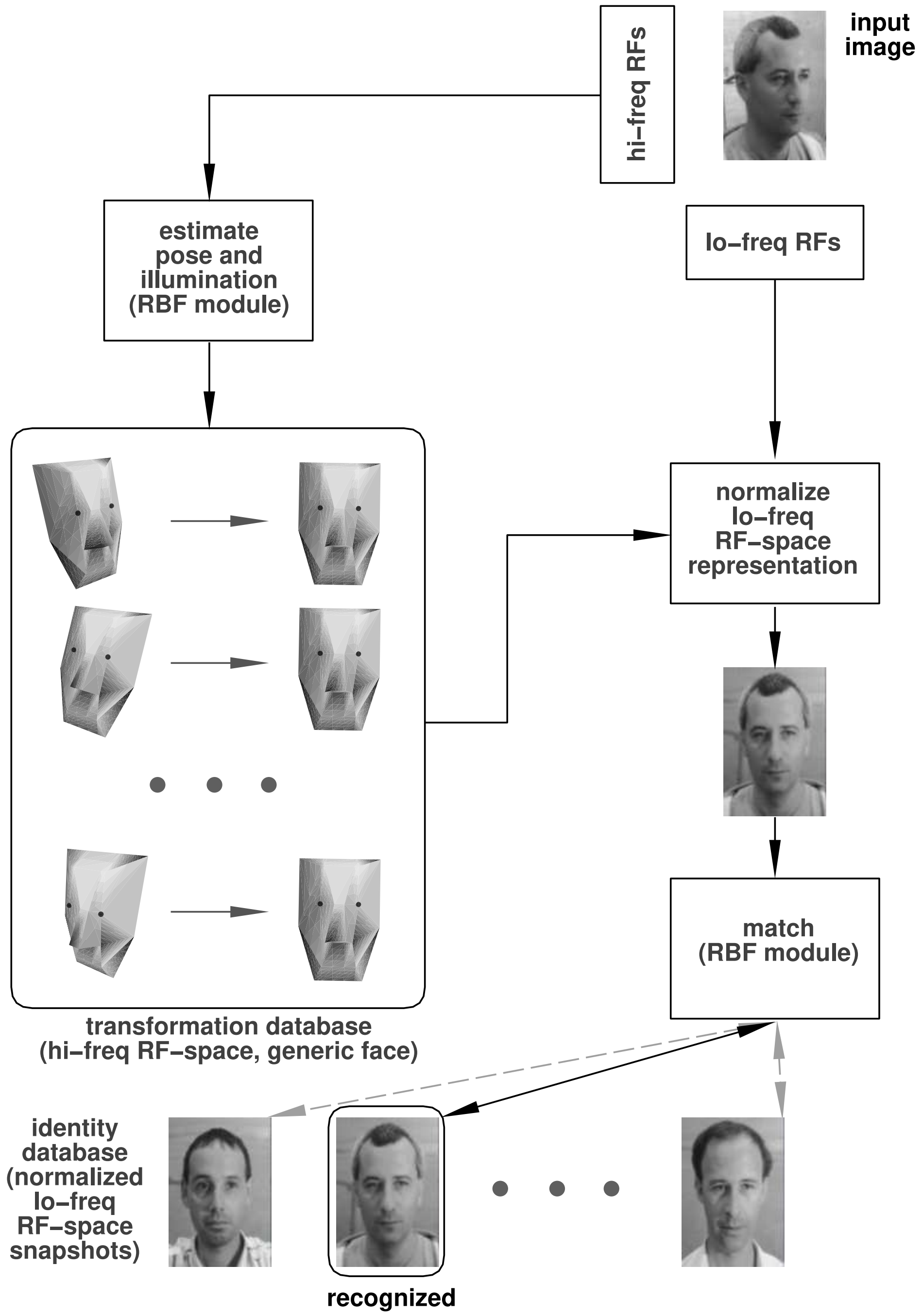


Figure 13: Block diagram of the complete recognition model (see section 4). Note that following the initial transduction by the low-frequency and the high-frequency RF modules, faces are represented throughout the system as vectors of activities of RFs, and not as the images which are included in this figure for illustration purposes.

faces, and identifies the input face.

In this model, the amount of "prior experience" with face transformations corresponds to the size of the set of individual face images, paired with the images of the same faces obtained under different but known viewing conditions. In our experiments, each face was represented by 15 images, taken under all combinations of 5 viewpoints and 3 illuminations. These conditions parallel closely the range of viewpoints and illuminations used in the psychophysical study of (Moses et al., 1993); only one value of illumination (corresponding to a superposition of two other illumination directions) was omitted.

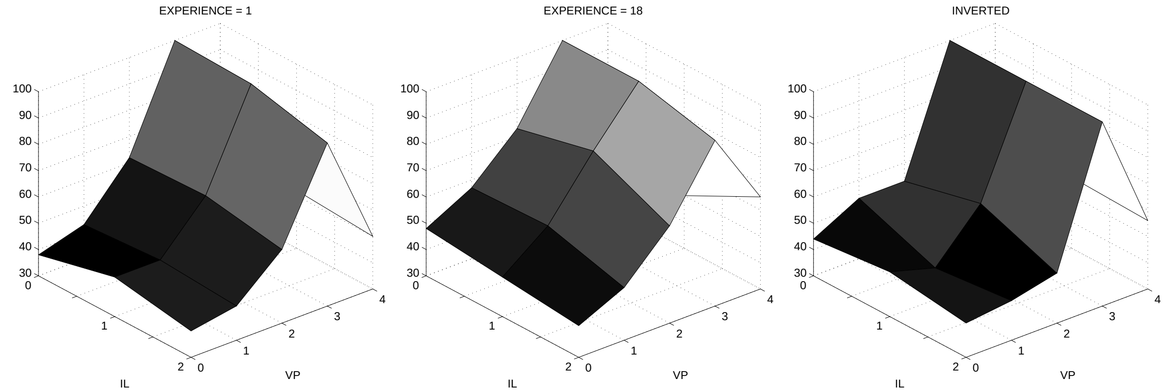


Figure 14: The improvement with experience in the model's generalization performance. *Left:* experience= 1, upright faces. *Middle:* experience= 18, upright faces. *Right:* experience= 18 with upright faces, testing with inverted faces (cf. (Moses et al., 1993)). The performance is similar to that of human subjects (Figure 15) in several respects (upright/inverted comparison; better generalization over illumination vs. viewpoint).

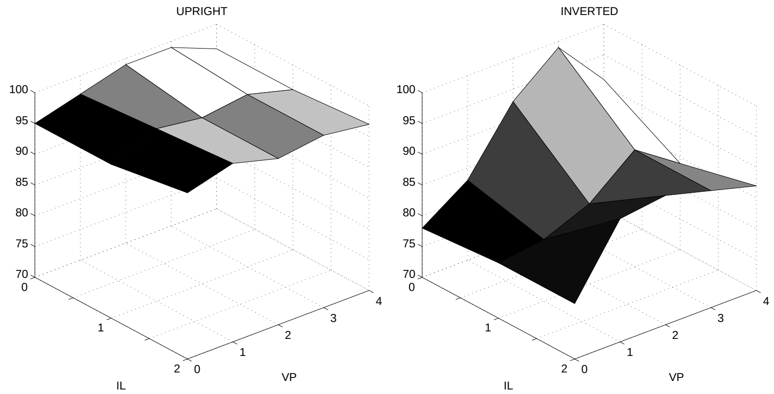


Figure 15: Human performance for upright (left) and inverted (right) face images (replotted from (Moses et al., 1993)).

4.1 Recovery of viewing conditions

We have trained an RBF classifier (Moody and Darken, 1989) to approximate the mapping from the high-frequency (16 *cpd*) RF-space representation ${}^{H}\mathbf{X}$ of a face image to the space of viewing conditions of the face. The viewing condition was encoded as a vector $\mathbf{V} \in R^{v}$, using unary representation (there were $v = 15$ possible viewing conditions). The n training images in the RF-space (${}^{H}\mathbf{X}^{(f_i)}_{\mathbf{V}_i}$, $i = [1 \ldots n]$), paired with the corresponding viewing conditions ($\mathbf{V}_i$), were used to estimate the parameters c_i and σ in the RBF approximation formula:

$$\mathbf{V}({}^{H}\mathbf{X}) = \sum_{i=1}^{n} c_i e^{-\frac{\| {}^{H}\mathbf{X} - {}^{H}\mathbf{X}^{(f_i)}_{\mathbf{V}_i} \|^2}{\sigma^2}} \tag{1}$$

which mapped an RF-space representation of an input image, ${}^{H}\mathbf{X}$, to the viewing condition vector $\mathbf{V}({}^{H}\mathbf{X})$ (the RBF centers were set to the training images ${}^{H}\mathbf{X}^{(f_i)}_{\mathbf{V}_i}$). The output of the module was defined to be the index of that element of $\mathbf{V}$ which was the closest to 1.

We explored the dependence of the performance of the resulting module on the number of training images available for each viewing condition. Performance was estimated from 10 trials, each involving 150 images, taken under viewing conditions different from those used in training (see Figure 16, left). It may be seen that the module is capable of quite accurate recovery of the viewing condition, after being trained on views of as few as 14 different faces.

4.2 Normalization to the prototypical view and identification

Once the viewing conditions $\mathbf{V}$ are determined by the view-recovery RBF classifier from the high-frequency RF-space representation, the proper class-specific transformation is applied in the low-frequency RF space, yielding the predicted prototype. The class-specific transformation here is calculated as the average of the transformations $\Delta {}^{L}\mathbf{X}^{(f_i)} = {}^{L}\mathbf{X}^{(f_i)}_{\mathbf{V}} - {}^{L}\mathbf{X}^{(f_i)}_{\mathbf{0}}$ taking each of the n different known faces from view $\mathbf{V}$ to the canonical view denoted by the subscript $\mathbf{0}$:

$${}^{L}\hat{\mathbf{X}}^{(f_0)}_{\mathbf{0}} = {}^{L}\mathbf{X}_{\mathbf{V}} - \frac{1}{n}\sum_{i=1}^{n} \Delta {}^{L}\mathbf{X}^{(f_i)} \tag{2}$$

The hypothesized prototype in its normalized form, ${}^{L}\hat{\mathbf{X}}^{(f_0)}_{\mathbf{0}}$, is passed on to the matching module, which, by interpolation among stored examples, identifies it with one of the familiar faces (for details regarding this procedure, see, e.g., Edelman et al., 1992). The identification performance is summarized in Figure 16, right, and in Table 1. Following each recognition trial, the particular transformation mapping the input view to its prototypical form is added to the system's database of

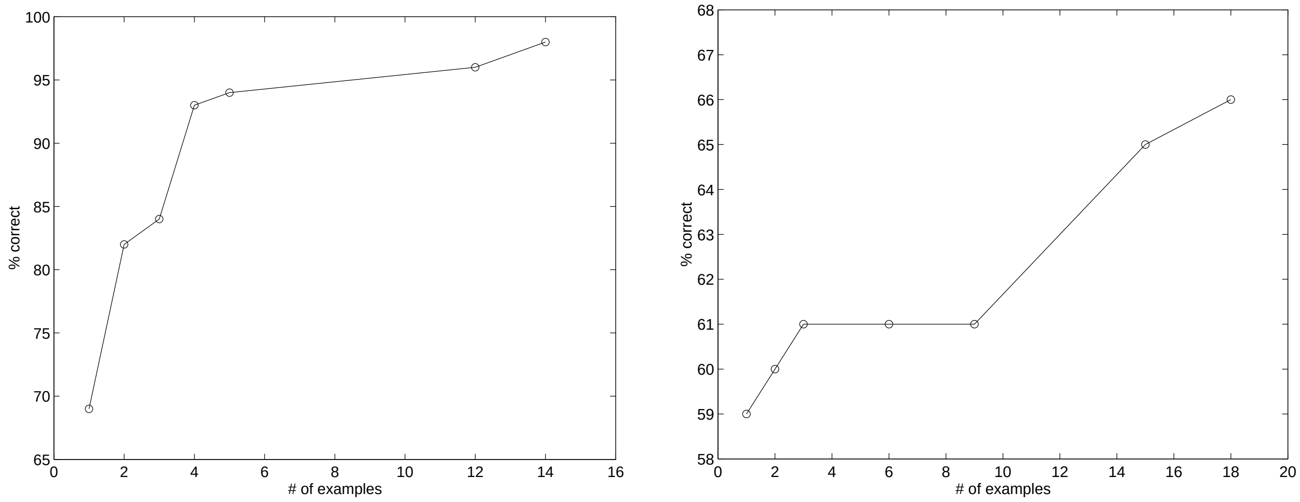


Figure 16: *Left:* performance in the determination of viewing condition, and its dependence on the extent of experience in seeing faces under varying viewing conditions (see section 4.1). The abscissa is the number of the images of each viewing condition used in training the RBF classifier. *Right:* the dependence of the generalization performance of the model (mean over all 15 viewing conditions) on experience. The abscissa shows the number of individuals whose images were used during the training stage.

class-specific transformations. In addition, the identity database (used by the matching module) can also be updated. In this manner, the model is capable of improving its performance with experience.

5 Comparing model and human performance

We compared the performance of the model with that of human subjects, by replicating the upright/inverted face generalization experiments described in (Moses et al., 1993). In the simulated experiment, the model was trained to recognize single images of three different faces taken under a fixed combination of viewpoint and illumination. The generalization performance of the model was then tested by computing the identification rate for images of the same faces under a wide range of viewpoints and illumination conditions. The difficulty of this task is noteworthy: the system was provided with just one view of each face, and had to recognize 15 other images of the same face, taken from viewpoints differing by rotation of up to 68^o around the vertical axis, and under widely varying illuminations.

The extent of the prior experience of the model was varied between 1 (seeing a single individual at the 15 available combinations of orientation and illumination) and 18 (seeing nearly all the indi-

V IL	-34	-17	0	17	34
0	38	43	62	100	55
1	45	45	63	99	52
2	40	43	58	92	50

V IL	-34	-17	0	17	34
0	48	57	73	100	58
1	45	58	80	100	52
2	42	50	67	93	65

Table 1: Generalization rate for test images, following exposure to a single training image (see also Figures 16, right, and 14). *Left:* experience= 1; *Right:* experience= 18. Note that chance-level performance in these three-alternative forced-choice experiments is 33% correct.

viduals in the database, under the 15 different viewing conditions). The model's mean generalization performance (correct recognition rate for novel views) grew from 59% at experience level 1 to 66% at experience level 18 (see Figure 16, right). Importantly, the increase in the level of experience on upright faces from 1 to 18 did not improve the model's generalization performance on inverted face images, replicating the main psychophysical finding of (Moses et al., 1993) (see Figure 14).

The experience-dependent increase in the mean generalization performance of the model from 59% to 66%, with the latter figure obtained with exposure to mere 18 individuals, indicates that further improvement is possible if the system is exposed to the wide range of face images normally seen by a human adult (hundreds of different faces, under a variety of viewing conditions). To parallel the essentially perfect ($\approx$ 97% correct; see Moses et al., 1993) generalization performance of human subjects for upright faces, certain computational sophistication may be required. One possible approach here is to rely on the symmetry property of faces: for bilaterally symmetric objects, a simple transformation of a generic 2D view of the object yields another legal view (Poggio and Vetter, 1992). For faces, this transformation corresponds to the mirroring of one view of a face with respect to the sagittal plane to obtain another view. The newly available view can then be used to improve the training of the viewpoint recovery module. When the knowledge of bilateral symmetry of faces was taken into account in our system in this manner, the generalization performance with experience= 18 was boosted from 66% to 76%. Further improvement in performance should be possible if the final classification is carried out not directly in the RF-space, but rather in a low-dimensional space spanned by an ensemble of individual-face recognition modules (cf. Edelman, 1995). The benefits of class-specific dimensionality reduction implemented by such a two-stage system include the improvement of about 20% in the recognition performance (Edelman et al., 1992); this approach should have a similar effect on the performance of the present model.

6 Summary and discussion

We have presented a model of the human ability to generalize face recognition from a single image. The model is based on a computational analysis of the notion of class-specific transformations, and is supported by computer simulations, in which it exhibited a considerable ability to generalize from single images. The model also replicated the central findings of a recent psychophysical study which examined generalization in upright and inverted faces in human subjects. From the practical standpoint, the model constitutes a significant advance over the approach of (Edelman et al., 1992), which also used RBF classifiers to learn face recognition from examples, but was incapable of generalization from a single image of a face. From the standpoint of theoretical neurobiology, the model suggests a promising approach to the understanding of the computational basis of class-based generalization in human vision.

The relevance of our model to the understanding of human vision stems not only from its performance in the upright/inverted face recognition experiments, but also from its predictions regarding the effect of face distinctiveness on generalization. The less similar a face is to an average human face (in a geometrical sense), the more different are the relationships among the RF-space representations of its views, compared to those of an average face. In such a case, one would tend to predict worse recognition results. However, the third step of the recognition in our model — comparing the predicted prototypical representation with those of known prototypes — is actually easier for a distinctive face, because the more unusual the face is, the larger the RF-space distance between its representation and those of the other faces. Consequently, it should be more difficult to generalize over novel views of distinctive faces, but also more difficult to misrecognize them under more familiar viewing conditions. These predictions are consistent with the results of recent experiments carried out by Newell et al. (1995), who found better performance for distinctive faces in the mismatch trials, but not in the match trials in a face matching experiment.

The interpretation offered by our model for the results obtained with distinctive faces can be extended to account for the peculiarities of recognition of faces across race (Brigham, 1986). It is well known that people used to see predominantly Caucasian faces find it more difficult to distinguish among Oriental faces than people living in the Orient, and vice versa. We conjecture that this happens because of the limited applicability of class-specific transformations to a radically different population of face shapes. With practice, the relevant portion of the representation space may become populated by the proper prototypes, and the discrimination performance may improve (as it indeed happens in cross-racial face recognition).

As a conclusion, we point out that the class-specific transformation approach adopted in the present work can be extended to classes of objects other than faces, and is currently under investigation in a wider context of 3D shape representation (Duvdevani-Bar and Edelman, 1995).

Acknowledgments. We thank Moshe Bar and Florin Cutzu for constructive advice and for help with Inventor graphics, and Ronen Basri, Yael Moses and Tomaso Poggio for useful discussions and for comments on an earlier version of this work. The database of controlled images of human faces used in this work is courtesy of Yael Moses, and is available by anonymous ftp at URL ftp://eris.wisdom.weizmann.ac.il/pub/FaceBase/.

References

Amari, S. (1968). Invariant structures of signal and feature spaces in pattern recognition problems. *RAAG Memoirs*, 4:553–566.

Amari, S. (1978). Feature spaces which admit and detect invariant signal transformations. In *Proc. 4th Intl. Conf. Pattern Recognition*, pages 452–456, Tokyo.

Basri, R. (1992). Recognition by prototypes. A.I. Memo No. 1391, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA. Intl. Journal of Computer Vision, in press.

Beymer, D., Shashua, A., and Poggio, T. (1993). Example based image analysis and synthesis. A.I. Memo No. 1431, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

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

Carey, S. and Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, 195:312–314.

Carey, S., Diamond, R., and Woods, B. (1980). Development of face recognition-a maturational component? *Developmental Psychology*, 16:257–269.

Diamond, R. and Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of experimental psychology*, 115(2):107–117.

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

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

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

Moody, J. and Darken, C. (1989). Fast learning in networks of locally tuned processing units. *Neural Computation*, 1:281–289.

Moses, Y., Edelman, S., and Ullman, S. (1993). Generalization across illumination and orientation changes for inverted and upright faces. CS-TR 14, Weizmann Institute of Science.

Newell, F., Chiroro, P., and Valentine, T. (1995). Recognising unfamiliar faces: The effects of distinctiveness and view. in preparation.

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

Poggio, T. and Vetter, T. (1992). Recognition and structure from one 2D model view: observations on prototypes, object classes, and symmetries. A.I. Memo No. 1347, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

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

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

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

Wilson, H. R. and Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research*, 19:19–32.

A Lemma

Lemma. The minimum of the expression

$$S = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2}} \tag{3}$$

where $\forall i, a_i(1-\delta) \le b_i \le a_i(1+\delta)$, $0 \le \delta \le a_i$ and $a_i \ne 0$, is equal to

$$\min(S)_{b_n} = \sqrt{1-\delta^2}$$

Proof. For every setting of $n-1$ terms in the above expression, the minimum of S will be achieved at one of the boundary points, with $b_n = a_n(1+\delta)$ or with $b_n = a_n(1-\delta)$. Without loss of generality, suppose that $n-1$ terms in expression 3 are already fixed.

Denote: $S_1 = \sum_{i=1}^{n-1} a_i b_i$, $\quad S_2 = \sum_{i=1}^{n-1} a_i^2$, $\quad S_3 = \sum_{i=1}^{n-1} b_i^2$. We can then write the sum in expression 3 for the entire n terms as follows:

$$S = \frac{S_1 + a_n b_n}{\sqrt{S_2 + a_n^2}\,\sqrt{S_3 + b_n^2}}$$

To find the value of b_n that minimizes this expression, we take its derivative with respect to b_n:

$$\frac{\partial S}{\partial b_n} = \frac{-S_1(b_n - a_n\frac{S_3}{S_1})}{(S_3 + b_n^2)^{3/2}(S_2 + a_n^2)^{1/2}}$$

The derivative is positive for b_n from $-\infty$ to $a_n\frac{S_3}{S_1}$ and negative for b_n from $a_n\frac{S_3}{S_1}$ to ∞. Depending on the particular values of S_1 and S_3, the minimum of S will be achieved at one of the boundary points, when $b_n = a_n(1+\delta)$ or when $b_n = a_n(1-\delta)$.

Let n_1 be the number of points where the minimization of S requires that $b_i = a_i(1+\delta)$, and n_2 — the number of points where it is required that $b_i = a_i(1-\delta)$, $n_1 + n_2 = n$. Then:

$$\begin{aligned}
S &= \frac{\sum_{i_1=1}^{n_1} a_{i_1}^2(1+\delta) + \sum_{i_2=1}^{n_2} a_{i_2}^2(1-\delta)}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i_1=1}^{n_1} a_{i_1}^2(1+\delta)^2 + \sum_{i_2=1}^{n_2} a_{i_2}^2(1-\delta)^2}} \\
&= \frac{1 + \delta\left(\frac{\sum_{i_1=1}^{n_1} a_{i_1}^2 - \sum_{i_2=1}^{n_2} a_{i_2}^2}{\sum_{i=1}^{n} a_i^2}\right)}{\sqrt{1 + \delta^2 + 2\delta\left(\frac{\sum_{i_1=1}^{n_1} a_{i_1}^2 - \sum_{i_2=1}^{n_2} a_{i_2}^2}{\sum_{i=1}^{n} a_i^2}\right)}}
\end{aligned}$$

Let $m = \left(\frac{\sum_{i_1=1}^{n_1} a_{i_1}^2 - \sum_{i_2=1}^{n_2} a_{i_2}^2}{\sum_{i=1}^{n} a_i^2}\right)$, $\quad -1 \le m \le 1$. Then

$$S = \frac{1 + \delta m}{\sqrt{1 + \delta^2 + 2\delta}}$$

Take the derivative to find the minimum of S:

$$\frac{\partial S}{\partial m} = \frac{\delta^2(m+\delta)}{(1 + \delta^2 + 2\delta m)^{3/2}}$$

The minimum of S will be achieved when $m = -\delta$ and will be equal to

$$\sqrt{1-\delta^2}$$

$\square$

B Similarity of RF activity changes

We now examine the hypothesis that changes of viewing conditions evoke similar changes of RF activity, regardless of the face identity. Suppose that we have images of two persons, $P1$ and $P2$. Consider two different viewing conditions (L,V) and (L',V'), where L and L' are different illumination directions, and V and V' — different viewing positions. Under each combination of viewing conditions, an image can be represented in the RF space as a vector of length n, equal to the number of RFs: $\mathbf{X}^{P1}_{(L,V)}$, $\mathbf{X}^{P1}_{(L',V')}$, $\mathbf{X}^{P2}_{(L,V)}$, $\mathbf{X}^{P2}_{(L',V')}$.

Let us compare the changes of RF activity precipitated by a change in the viewing conditions: $\mathbf{\Delta X}^{P1} = \mathbf{X}^{P1}_{(L,V)} - \mathbf{X}^{P1}_{(L',V')}$, and $\mathbf{\Delta X}^{P2} = \mathbf{X}^{P2}_{(L,V)} - \mathbf{X}^{P2}_{(L',V')}$. We decompose the transformation $(L,V) \to (L',V')$ into $(L,V) \to (L',V)$ and $(L',V) \to (L',V')$, and study the changes of activity in each of these two cases (illumination direction change and viewing position change) separately. We assume that the specular component of reflection function is small and take into account only the diffuse component. To represent relative similarity of shape of different faces, we introduce a parameter δ and assume that each coordinate of the normal at each particular point of the face can vary from person to person at most by δ.

B.1 Case 1: $(L,V) \to (L',V)$

We assume that all RFs have the same profile and that they are small enough, so that intensity at each point of the image under each particular RF is constant. In this case the result of the convolution will be a multiplication of that intensity value by a constant k, which depends on the RF profile and will be omitted from further calculations. Let $I(i)_{(L,V)}$ be the intensity of the image patch corresponding to the i'th RF, under the viewing conditions (L,V). Then the RF activities and their difference are:

$$\begin{aligned}
\mathbf{X}_{(L,V)} &= \left[I(1)_{(L,V)},\ I(2)_{(L,V)},\ \ldots,\ I(n)_{(L,V)}\right] \\
\mathbf{X}_{(L',V)} &= \left[I(1)_{(L',V)},\ I(2)_{(L',V)},\ \ldots,\ I(n)_{(L',V)}\right] \\
\mathbf{\Delta X} &= \left[(I(1)_{(L,V)} - I(1)_{(L',V)}),\ \ldots,\ (I(n)_{(L,V)} - I(n)_{(L',V)})\right]
\end{aligned}$$

In the above expressions, the intensity values for persons $P1$ and $P2$ are

$$
\begin{aligned}
I(i)_{(L,V)} &= (L \cdot N_i^1) \\
I(i)_{(L',V)} &= (L' \cdot N_i^1) \\
I(i)_{(L,V)} &= (L \cdot N_i^2) \\
I(i)_{(L',V)} &= (L' \cdot N_i^2)
\end{aligned}
$$

The changes in RF activities for the two persons are:

$$
\begin{aligned}
\mathbf{\Delta X}_{P1} &= \left[\left((L-L') \cdot N_1^1\right), \left((L-L') \cdot N_2^1\right), \ldots, \left((L-L') \cdot N_n^1\right)\right] \\
\mathbf{\Delta X}_{P2} &= \left[\left((L-L') \cdot N_1^2\right), \left((L-L') \cdot N_2^2\right), \ldots, \left((L-L') \cdot N_n^2\right)\right]
\end{aligned}
$$

We now compare the lengths and directions of these RF activity vectors, assuming that corresponding RFs for different persons will be positioned over regions at which the normals can differ at most by δ in each component:

$$
\begin{aligned}
\|\mathbf{\Delta X}_{P1}\|^2 &= \sum_{i=1}^{n} \left((L-L') \cdot N_i^1\right)^2 \\
\|\mathbf{\Delta X}_{P2}\|^2 &= \sum_{i=1}^{n} \left((L-L') \cdot N_i^2\right)^2 \\
&\leq \sum_{i=1}^{N} \left((L-L') \cdot (N_i^1 + \delta N_i^1)\right)^2 \\
&= (1+\delta)^2 \|\mathbf{\Delta X}_{P1}\|^2 \\
\\
\|\mathbf{\Delta X}_{P2}\|^2 &\geq \sum_{i=1}^{n} \left((L-L') \cdot (N_i^1 - \delta N_i^1)\right)^2 \\
&= (1-\delta)^2 \|\mathbf{\Delta X}_{P1}\|^2
\end{aligned}
$$

yielding

$$
\cos \angle(\mathbf{\Delta X}_{P1}\ ,\ \mathbf{\Delta X}_{P2}) = \frac{\sum_{i=1}^{n} \left((L-L') \cdot N_i^1\right)\left((L-L') \cdot N_i^2\right)}{\|\mathbf{\Delta X}_{P1}\| \|\mathbf{\Delta X}_{P2}\|} \tag{4}
$$

Let $a_i = \left((L-L') \cdot N_i^1\right)$ and $b_i = \left((L-L') \cdot N_i^2\right)$. Then

$$
a_i(1+\delta) \leq b_i \leq a_i(1+\delta)
$$

and, according to the lemma of section A, the minimum value of the expression $\frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2}}$ is $\sqrt{1-\delta^2}$. Therefore:

$$\cos \angle(\mathbf{\Delta X}_{P1}\ ,\ \mathbf{\Delta X}_{P2})\ \geq \sqrt{1-\delta^2}$$

B.2 Case 2: $(L', V) \rightarrow (L', V')$

As before,

$$\begin{aligned}
\mathbf{X}_{(L',V)} &= \left[I(1)_{(L',V)},\ I(2)_{(L',V)},\ \ldots,\ I(n)_{(L',V)}\right] \\
\mathbf{X}_{(L',V')} &= \left[I(1)_{(L',V')},\ I(2)_{(L',V')}\ \ldots,\ I(n)_{(L',V')}\right]
\end{aligned}$$

where the intensity values for images of persons $P1$ and $P2$ are:

$$\begin{aligned}
I(i)_{(L',V)} &= (L' \cdot N_i^1) \\
I(i)_{(L',V')} &= (L' \cdot N_i^{1'}) \\
I(i)_{(L',V)} &= (L' \cdot N_i^2) \\
I(i)_{(L',V')} &= (L' \cdot N_i^{2'})
\end{aligned}$$

As the viewing position changes, the RFs move over the image. The changes of the normal to the surface over different regions of the object will then cause change in RF response:

$$\begin{aligned}
\mathbf{\Delta X}_{P1} &= [(L' \cdot (N_1^1 - N_1^{1'})),\ (L' \cdot (N_2^1 - N_2^{1'})),\ \ldots, (L' \cdot ({N^1}_N - N_n^{1'}))] \\
\mathbf{\Delta X}_{P2} &= [(L' \cdot (N_1^2 - N_1^{2'})),\ (L' \cdot (N_2^2 - N_2^{2'})),\ \ldots,\ (L' \cdot (N_n^2 - N_n^{2'}))]
\end{aligned}$$

This gives

$$\begin{aligned}
\|\mathbf{\Delta X}_{P1}\|^2 &= \sum_{i=1}^{n} \left(L' \cdot (N_i^1 - N_i^{1'})\right)^2 \\
\|\mathbf{\Delta X}_{P2}\|^2 &= \sum_{i=1}^{n} \left(L' \cdot (N_i^2 - N_i^{2'})\right)^2
\end{aligned}$$

Let $\Delta N_i = N_i^2 - N_i^1$ and $\Delta N_i' = N_i^{2'} - N_i^{1'}$. We proceed to estimate the upper bound of the $\|\mathbf{\Delta X}_{P2}\|^2$ in terms of $\|\mathbf{\Delta X}_{P1}\|^2$.

$$\begin{aligned}
\|\mathbf{\Delta X}_{P2}\|^2 &= \sum_{i=1}^{n}\left(L'\cdot\left(N_i^1 - N_i^{1'} + \Delta N_i - \Delta N_i'\right)\right)^2 \\
&= \sum_{i=1}^{n}\left(\left(L'\cdot(N_i^1 - N_i^{1'})\right) + (L'\cdot\Delta N_i) - (L'\cdot\Delta N_i')\right)^2 \\
&\leq \sum_{i=1}^{n}\left(\left|\left(L'\cdot(N_i^1 - N_i^{1'})\right)\right| + \max\left|(L'\cdot\Delta N_i) - (L'\cdot\Delta N_i')\right|\right)^2
\end{aligned} \tag{5}$$

Now, if $(L'\cdot N_i^1) < (L'\cdot N_i^{1'})$ then

$$\max\left|(L'\cdot\Delta N_i) - (L'\cdot\Delta N_i')\right| = \max(L'\cdot\Delta N_i) - \min(L'\cdot\Delta N_i')$$

and if $(L'\cdot N_i^1) \geq (L'\cdot N_i^{1'})$ then

$$\max\left|(L'\cdot\Delta N_i) - (L'\cdot\Delta N_i')\right| = \min(L'\cdot\Delta N_i) - \max(L'\cdot\Delta N_i')$$

Note that

$$\max(L'\cdot\Delta N_i) = \max\|L'\|\|\Delta N_i\|\cos\angle(L,\Delta N_i)$$

and

$$\max\|\Delta N_i\| = \sqrt{(\delta N1_{i_x})^2 + (\delta N1_{i_y})^2 + (\delta N1_{i_z})^2} = \delta$$

Therefore:

$$\max(L'\cdot\Delta N_i) \leq \delta$$

and, similarly, one can show that

$$\max(L'\cdot\Delta N_i') \leq \delta$$

The lower bounds are, respectively:

$$\min(L'\cdot\Delta N_i) \geq -\delta$$

and

$$\min(L'\cdot\Delta N_i') \geq -\delta$$

Equation 5 then yields:

$$\begin{aligned}
\|\mathbf{\Delta X}_{P2}\|^2 &\leq \sum_{i=1}^{n}\left(\left|(L'\cdot(N_i^1 - N_i^{1'}))\right| + 2\delta\right)^2 \\
&= \sum_{i=1}^{n}\left[\left|(L'\cdot(N_i^1 - N_i^{1'}))\right|\left(1 + \frac{2\delta}{|(L'\cdot(N_i^1 - N_i^{1'}))|}\right)\right]^2
\end{aligned}$$

where $\left|(L'\cdot(N_i^1 - N_i^{1'}))\right|$ is the modulus of the variation in the projection of some normal onto the illumination direction (this depends on the location of the RF on the image and on the changes in the viewing position).

The geometry of a face is very complex: for any variation in viewing position, normals at the different points of the image can change in very dissimilar ways. However, we can bound this value from below. It cannot be equal to 0 at all the points, for then there would be no change in the RF activities. Denote the minimum non-zero value that this expression can attain by $(L'\cdot(\epsilon\mathbf{N}))$. Then

$$\begin{aligned}
\|\mathbf{\Delta X}_{P2}\|^2 &\leq \sum_{i=1}^{n}\left[\left|(L'\cdot(N_i^1 - N_i^{1'}))\right|\left(1 + \frac{2\delta}{(L'\cdot(\epsilon\mathbf{N}))}\right)\right]^2 \\
&\leq \left(1 + \frac{2\delta}{(L'\cdot(\epsilon\mathbf{N}))}\right)^2 \|\mathbf{\Delta X}_{P1}\|^2
\end{aligned}$$

Similarly, we can obtain the lower bound of $\|\mathbf{\Delta X}_{P2}\|^2$ in terms of $\|\mathbf{\Delta X}_{P1}\|^2$, and, as a result, estimate the bound on the angle between the two vectors, $\mathbf{\Delta X}_{P1}$ and $\mathbf{\Delta X}_{P2}$:

$$\begin{aligned}
\|\mathbf{\Delta X}^{P2}\|^2 &\geq \left[\sum_{i=1}^{n}\left(\left|\left(L'\cdot(N_i^1 - N_i^{1'})\right)\right| + \min\left|(L'\cdot\Delta N_i) - (L'\cdot\Delta N_i')\right|\right)^2\right] \\
\|\mathbf{\Delta X}_{P2}\|^2 &\geq \left(1 - \frac{2\delta}{(L'\cdot(\epsilon\mathbf{N}))}\right)^2 \|\mathbf{\Delta X}_{P1}\|^2 \\
\cos\angle(\mathbf{\Delta X}_{P1}, \mathbf{\Delta X}_{P2}) &= \frac{\sum_{i=1}^{n}(L'\cdot(N_i^1 - N_i^{1'}))(L'\cdot(N_i^2 - N_i^{2'}))}{\|\mathbf{\Delta X}_{P1}\|\|\mathbf{\Delta X}_{P2}\|}
\end{aligned}$$

Let $a_i = (L'\cdot(N_i^1 - N_i^{1'}))$ and $b_i = (L'\cdot(N_i^2 - N_i^{2'}))$. Then

$$a_i\left(1 - \frac{2\delta}{(L'\cdot(\epsilon\mathbf{N}))}\right) \leq b_i \leq a_i\left(1 + \frac{2\delta}{(L'\cdot(\epsilon\mathbf{N}))}\right)$$

Then, as in the case of changes in illumination direction, we can use the Lemma. The lower bound on the cosine of the angle between $\mathbf{\Delta X}_{P1}$, and $\mathbf{\Delta X}_{P1}$ will be:

$$
\begin{aligned}
\cos\angle(\boldsymbol{\Delta}\mathbf{X}_{P1}, \boldsymbol{\Delta}\mathbf{X}_{P2}) &= \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2}} \\
&\geq \sqrt{1 - \left(\frac{2\delta}{(L' \cdot (\epsilon\mathbf{N}))}\right)^2}
\end{aligned}
$$