

Face Recognition: The Problem of Compensating for Changes in Illumination Direction

Yael Adini, Yael Moses, and Shimon Ullman

Abstract—A face recognition system must recognize a face from a novel image despite the variations between images of the same face. A common approach to overcoming image variations because of changes in the illumination conditions is to use image representations that are relatively insensitive to these variations. Examples of such representations are edge maps, image intensity derivatives, and images convolved with 2D Gabor-like filters. Here we present an empirical study that evaluates the sensitivity of these representations to changes in illumination, as well as viewpoint and facial expression. Our findings indicated that none of the representations considered is sufficient by itself to overcome image variations because of a change in the direction of illumination. Similar results were obtained for changes due to viewpoint and expression. Image representations that emphasized the horizontal features were found to be less sensitive to changes in the direction of illumination. However, systems based only on such representations failed to recognize up to 20 percent of the faces in our database. Humans performed considerably better under the same conditions. We discuss possible reasons for this superiority and alternative methods for overcoming illumination effects in recognition.

Index Terms—Face recognition, compensating for illumination, edge representation, 2D Gabor-like filter, image comparison.

1 INTRODUCTION

FACE recognition is a difficult problem because of the generally similar shape of faces combined with the numerous variations between images of the same face. The image of a face changes with facial expression, age, viewpoint, illumination conditions, noise, etc. The task of a face recognition system is to recognize a face in a manner that is as independent as possible of these image variations. Psychophysical experiments show that the human visual system can identify faces from novel images despite considerable variations between images of the same face that are due to changes in illumination [28] and viewpoint [31], [11], [2], [28]. The question, then, is how a recognition system can identify a face despite these variations. Here we focus mainly on variations that are due to changes in illumination. We present an empirical study that evaluates commonly used approaches to overcoming image variations because of these changes. We also evaluate how these approaches affect variations between images of the same face because of changes in viewpoint and expression.

Three main approaches for dealing with image variations that are due to illumination changes have been used in the past. These approaches are used by general object recognition systems as well as by systems that are specific

to faces. The first approach uses the gray-level information to extract the three-dimensional shape of the object, namely, a shape from shading approach (e.g., [21]). This is an ill-posed problem and, therefore, all proposed solutions assume either the object shape and reflectance properties or the illumination conditions. These assumptions are too strict for general object recognition, and therefore were not shown to be sufficient for the face recognition task.

The second approach, which will be studied here, is based on representations of the image and the stored model that are relatively insensitive to changes in illumination. For example, the edge map of the image [12], [26], [20], [7], [40], [14] is often considered as the basic image representation model for general object recognition and, in particular, for face recognition [22], [47], [17], [3]. Other examples of image representations will be considered below.

The third approach to handle image variations that are due to illumination differences is by using as a model several images of the same object (face) taken under different illumination conditions. Here, the images can be used either as independent models or combined into a model-based recognition system [15], [19], [41]. Such approaches will be discussed in Section 4.

Ideally, an image representation used for recognition system should be invariant to illumination changes. It has been shown theoretically that, for the general case, a function invariant to illumination does not exist [29].¹ The objects considered by Moses and Ullman were unconstrained 3D objects, consisting of n independent patches in space. For recognition systems that are limited to certain classes of

• Y. Adini is with the Neurobiology Department, Brain Research, The Weizmann Institute of Science, Rehovot, 76100, Israel.
E-mail: adini@nisan.weizmann.ac.il.

• Y. Moses and S. Ullman are with the Computer Science Department, The Weizmann Institute of Science, Rehovot, 76100, Israel.
E-mail: {yael, shimon}@wisdom.weizmann.ac.il.

Manuscript received 22 May 1995. Recommended for acceptance by J. Daugman.
For information on obtaining reprints of this article, please send e-mail to: transpami@computer.org, and reference IEEECS Log Number 104915.

1. Similar results regarding variations that are due to changes of viewpoint were shown by [5], [8], [29].

objects, this limitation does not necessarily apply. Indeed, edge maps can serve as relatively robust representations of illumination changes for some classes of objects, such as telephones, tables, etc. However, for other objects, such as faces, part of the edges do not remain stable.² It remains an open question whether edge maps and other possible representations provide an illumination-insensitive representation for face recognition.

In the past, no qualitative or quantitative study was performed that addressed the effect of the imaging conditions, such as viewpoint and illumination, on the variations between images of the same objects. Such a systematic study is important to gain a better understanding of the actual problem that recognition systems must solve. The question addressed in this paper is whether several widely used image representations (such as edge maps) that are often considered to be insensitive to illumination changes, are sufficient for recognizing faces under different illumination. To answer this question, an empirical study was performed that evaluated the sensitivity of several commonly used image representations to changes in viewing conditions (e.g., viewpoint and illumination) when a face recognition task is considered. Here we used a special database of faces, in which each of the imaging conditions (illumination, viewpoint, and expression) was separately controlled. The distances between the pairs of images (or image representations) of different individuals (taken under the same conditions) were computed and compared with the distances between the pairs of images of the same face that varied because of a change in the viewing condition (illumination, viewpoint, and expression). The database was constructed so that the performance of recognition methods could be evaluated with respect to a single imaging parameter (e.g., illumination). We used for each imaging parameter only a single change in direction, i.e., a horizontal change of illumination or viewing direction. Our database can be used to study the limitations of compensating approaches to recognition. However, for demonstrating good performance of a recognition system, it is necessary to evaluate it with images that contain more complex variations, such as different changes in illumination and viewpoint directions, and additional transformations, changes of scale, and change of background. Furthermore, since natural variations between images are due to more than a single parameter, a recognition system should also be evaluated with images that vary because of a combination of imaging parameters.

The following image representations were considered: the original gray-level image, the edge map of the image, the image filtered with 2D Gabor-like filters, and the first and second derivatives of the gray-level image. Each of these image representations was constructed with several different parameter settings of the operators in question. We next briefly discuss the image representations under investigation.

1.1 Edge Map

Intensity edges coincide, generally, with gray-level transitions. Gray-level transitions can be due to discontinuities in

the surface color (albedo) or orientation. Such edges are expected to be insensitive to illumination changes. Other edges in the image may be related to illumination changes, including shadows and specularities. The advantage of using an edge representation is that it is a relatively compact representation (compared with the full gray-level image) and it is often insensitive to illumination changes for a variety of objects. Such edge representations were used by several face recognition systems [22], [47], [17], [3].

1.2 The Image Filtered with 2D Gabor-Like Functions

Physiological and psychophysical evidence indicates that at the early stages of human visual processing the images are processed by local, multiple, and parallel channels that are sensitive to both spatial frequency and orientation. Psychophysical evidence for the existence of such channels comes mainly from studies that use

- 1) summation at the threshold [6], [18],
- 2) selective adaptation [1], and
- 3) masking paradigms [39].

Physiological studies found cells in V1 (simple cells) that are selectively tuned to orientation as well as to spatial frequency and phase. It was suggested that the response profile of a simple cell could be approximated by 2D Gabor-like filters [9], [10], [13], [25], [33], [34] or a set of m th ordered spatial derivatives of a Gaussian [23]. Inspired by these findings, several artificial face recognition systems filter the gray-level image by a set of 2D Gabor-like functions before attempting to recognize the faces in the image [3], [4], [24]. Note that convolving the image with 2D Gabor-like functions is often similar to enhancing edge contours, as well as valleys and ridge contours from the image.

1.3 Derivatives of the Gray-Level

Derivatives of the gray-level distribution were used by several face recognition systems [3], [15] to reduce the effects of changes in illumination conditions on face images. The derivatives used include directional and nondirectional first and second order derivatives. It can be shown analytically that, under certain conditions, changes in the ambient light will indeed affect the gray-level image but not its derivatives. However, this is not the case in the natural conditions where the direction of the light source is also changed.

1.4 Log Transformations

In addition to the above representations, a nonlinear transformation often used in computer vision is the logarithmic transformation of the image intensities [35]. There is also physiological evidence that the response of cells in the retina is nonlinear in the intensity of the incoming image, which can be approximated as a log function of the intensity [46].

This paper is organized as follows: Section 2 describes the study performed, Section 3 presents its results, and Section 4 summarizes and discusses the study and its implications to the design of recognition systems that can deal with image variations because of changes in illumination. We have also considered the results established here relative to the performance of the human visual system in compensating for changes between images of the same face that are due to changes in illumination, viewpoint, and expression.

2. It can be shown theoretically that edges on a smooth surface are not stable with changes in the direction of illumination [27].



Fig. 1. One image of each of the 25 faces: frontal view with left illumination.

2 METHODS

Five images of each of 25 faces, taken from a well-controlled database of faces, were considered (see also Appendix A.1). All faces were of males without distinctive features such as glasses, beards, or mustaches (see Fig. 1). All images were taken by the same camera under tightly controlled conditions of illumination and viewpoint. Normalized frontal views for all faces were obtained by fixing the location of the face symmetry axis, the external corners of the eyes, and the bottom of the nose, before the pictures were taken (see Fig. 2). The following five images of each face were considered:

- 1) frontal view, left illumination and neutral expression;
- 2) frontal view, right illumination and neutral expression;

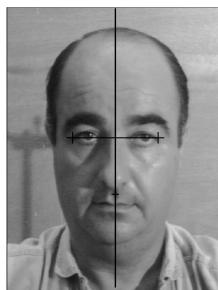


Fig. 2. Each face was normalized before taking the picture so that the face's symmetry axis, the external corners of the eyes, and the bottom of the nose were located on the reference lines as shown.

- 3) 34° rotation of the face around the vertical axis, left illumination and neutral expression;
- 4) frontal view, left illumination and a smile;
- 5) frontal view, left illumination, with closed eyes and open mouth (see Fig. 3).

The distances between all pairs of images of different faces taken under the same conditions (similar to each pair of images in Fig. 1) were measured and compared to the distances between the pairs of images of the same face that can vary because of



Fig. 3. An example of five images of the same face. (a) Frontal view with left illumination. (b) Frontal view with right illumination. (c) 34° away from the frontal view on the horizontal axis with left illumination. (d) A smile expression taken from frontal view with left illumination. (e) A drastic expression taken from the frontal view left illumination.

- 1) light source position: left vs. right, (Fig. 3a vs. Fig. 3b),
- 2) viewpoint: frontal vs. 34° to the left view, (Fig. 3a vs. Fig. 3c), and
- 3) expression: neutral vs. smile or neutral vs. drastic expression (Fig. 3a vs. Fig. 3d or Fig. 3e).

As previously mentioned, we considered a variety of image representations and distance measures that had been used for image comparison, in particular, edge maps, derivatives of the gray-level, the images filtered with 2D Gabor-like functions, and a representations that combined a Log function of the intensity with these representations. We considered different parameters of the operators that produced each of these representations (see Appendix A3). For example, for the 2D Gabor-like filters, the effects of the orientation, scale, and symmetry of the filters were studied, and, for the derivatives of the gray-level, the effects of the orientation, symmetry, and scale were studied. All together, 107 operators were considered. The distance between the pairs of images was computed using five simple distance measures between general 2D distributions (see Section A.2). These measures indicate that the objective distance between the images is not based on knowledge about the image formation function.

To avoid background interference, we extracted and considered only the face part of the image. Some psychological evidence suggests that different face parts make dissimilar contributions to face recognition (see a review by Shepherd et al. [38]). Therefore, we considered separately several regions of the face: the entire face without the hair, the eyes, and the lower part of the face (see Fig. 4).

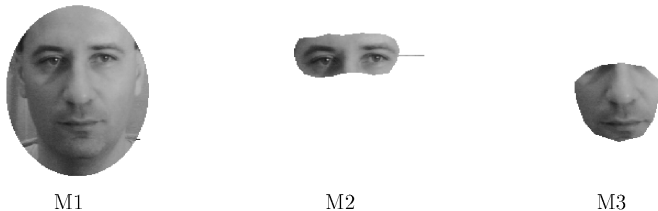


Fig. 4. The three masks of the face. (M1) The entire face (without the hair). (M2) The eyes area, and (M3) the lower part of the face.

The crucial question we were interested in answering was, for any of these representations, whether the distances between different images of the same individual would tend to be smaller than the distance between different individuals. Such representations could serve a useful role in compensating for image variations caused by changes in the viewing condition in object recognition.

3 RESULTS

In this section, we present the results of comparing the variation between pairs of images of different faces with the variation between different images of the same face. The variations between images of the same face were due to illumination (left vs. right), viewpoint (frontal vs. 34°), and expression (neutral vs. smile or vs. drastic expression). All together, 107 operators were considered, for 25 different

faces. In total, we have compared five different distance measures of about 100,000 pairs of images for three different face masks.

Ideally, to use some distance measures for the purpose of recognition, we would like the distances between images of different faces to be larger than the distances between images of the same face. None of the representations considered here had this property. However, some of the representations were found to compensate better than others for image variations. To make quantitative comparisons, we used a relative score to evaluate the sensitivity of a given representation to the variations between images of the same face relative to the variations between images of different faces. This was obtained by testing the performance of a given representation and distance measure with respect to our database.

For a particular representation and distance measure that we wanted to evaluate, we classified a given face as a *missed-face* if the system failed to recognize it. That is, a face was defined as missed-face if the distance between two images of this face taken under different conditions (e.g., different illumination) was larger than the distance between the image of the face and one of the images of another face (taken under the same conditions). The percentage of missed faces from the set of faces in our database, which we denoted by *miss-percentage*, was used to evaluate a given image representation with respect to a given imaging parameter. Zero and 100 percent miss-percentages correspond to perfect recognition and total failure of the system, respectively. The system may fail to recognize a given face because it confuses it with one or more faces from its database. Let the *failure-rate* be the average percentage of faces that the system confuses for each missed face. If the failure-rate is high, then it is unlikely that the system would misidentify a face merely because of similar arbitrary pairs of faces in the database.

3.1 Illumination Direction

We begin by analyzing the variations between the images that are due to a change in the illumination direction, that is, when images of the same face vary only because of the illumination direction, left vs. right. We will first consider a recognition system that is based on computing the distance between unprocessed gray-level images. We found that such a system will fail to recognize all the faces in the database—the miss-percentage is 100 percent. Furthermore, such a system will confuse each face with all the other faces in the database—the failure-rate is 100 percent. We have concluded that when comparing unprocessed images, the changes induced by illumination are larger than the differences between individuals.

We now consider the variations between processed images that are due to a change in the illumination direction compared to variations between different individuals taken under the same illumination. In particular, we studied the effects of the 107 different representations of the image variations. Fig. 5 shows the histogram of the miss-percent and failure-rate of all the operators and the parameters considered. This histogram is for the entire face mask with the best distance measure. The miss-percent and the failure-rate

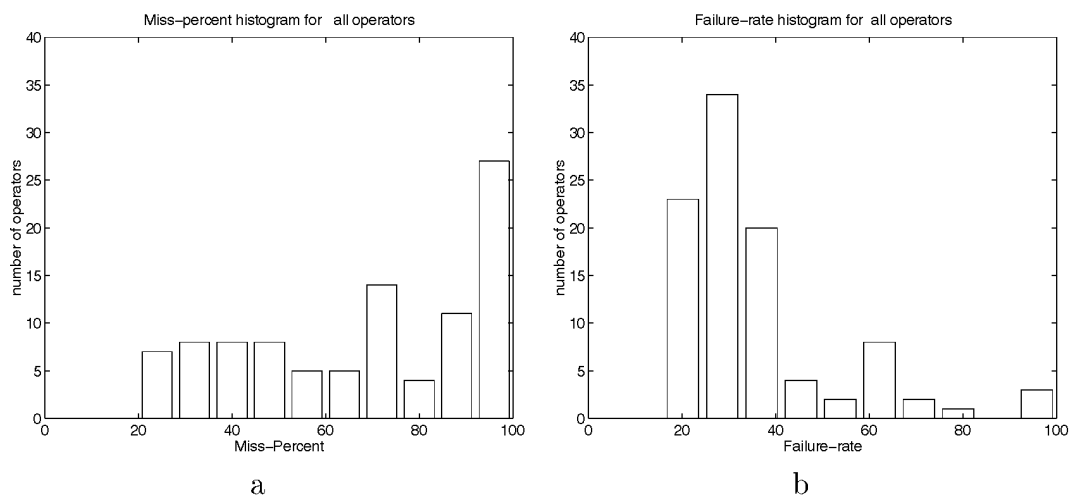


Fig. 5. (a) Histogram of the miss percent of all 107 operators. (b) Histogram of the failure rate of all 107 operators. Both histograms are for local affine-GL distance.

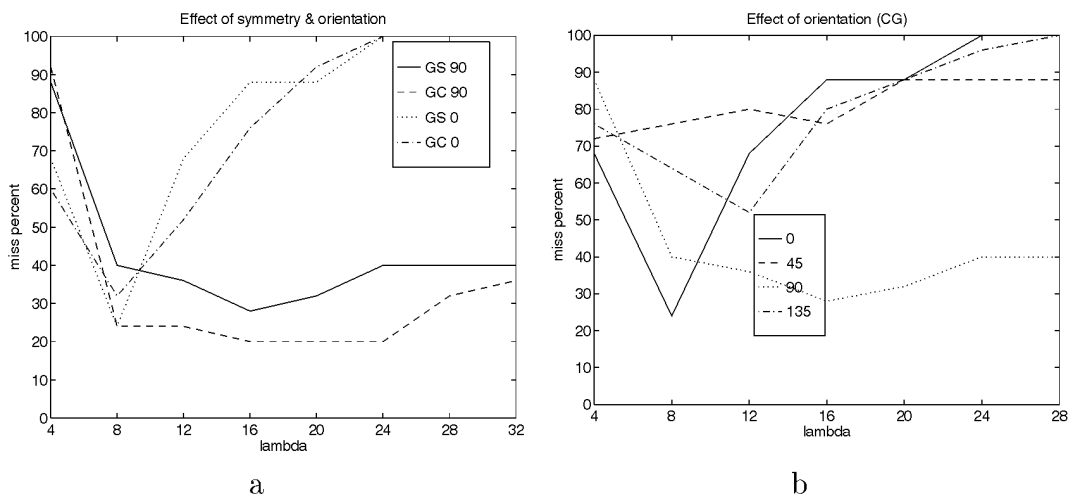


Fig. 6. (a) The effect of the Gabor symmetry (sine vs. cosine) and orientation (90° vs. 0° relative to the vertical axis) for different Gabor scales. The distance measure used is local affine GL. (b) The effect of the orientation for the Gabor cosine for different Gabor scale. The distance measure used is local affine GL.

vary between 20 percent and 100 percent. The miss-percent for most of the operators considered was above 50 percent.

We next consider which of the operator parameters reduces the miss-percent. We will consider the orientation, the scale,³ and the symmetry (e.g., cosine vs. sine) of the operator. The operator orientation strongly affects the miss-percent. This effect depends on the scale of the operator. For the larger scales, the miss-percent of a 2D Gabor-like filter which is sensitive to horizontal features, is reduced to 20 percent compared to over 60 percent for the other orientations (0° , 45° , and 135° from the vertical axis, Fig. 6b). This effect may be due to the horizontal orientation of facial features (e.g., mouth and eyes). A similar orientation effect was also found when the image was convolved with directional derivatives of a Gaussian. We studied 10 different

operator scales between $\sigma = 2$ to 20, which corresponds to $\lambda = 4$ to 40 (about 40 to four cycles per image). The *scale* of the operators was found to have a strong effect on the miss-percent. Fig. 6 illustrates this with respect to the 2D Gabor-like operator. The scales that gave the lowest miss-percent (20 percent) were from $\lambda = 12$ to 24. The miss-percent increased for larger scales, but despite the substantial smoothing of the faces in these representations (see Fig. 6a) it was still relatively low (36 percent), which will be discussed later. A similar effect was found for the scale of the Gaussian that was used to smooth the image before computing the directional derivatives, the Laplacian, and the edge detector. The *symmetry* of the 2D Gabor-like operator (i.e., sine or cosine) had almost no effect on the miss-percent (see Fig. 6a). The failure-rate of the operators with the lowest miss-percent was above 40 percent. In fact, even for the best operators, the system will fail to recognize 20 percent of the faces in our database. This failure does not arise from an accidental similarity between a single pair of faces, but

3. The standard deviations (σ) and the wave-length (λ) were chosen such that $\sigma = \lambda/2$; the size of the mask was chosen to be 2λ . Therefore, λ uniquely determined the scale of the operator. The units of λ were taken in pixels, where the size of the face in the image was about 170×170 pixels.

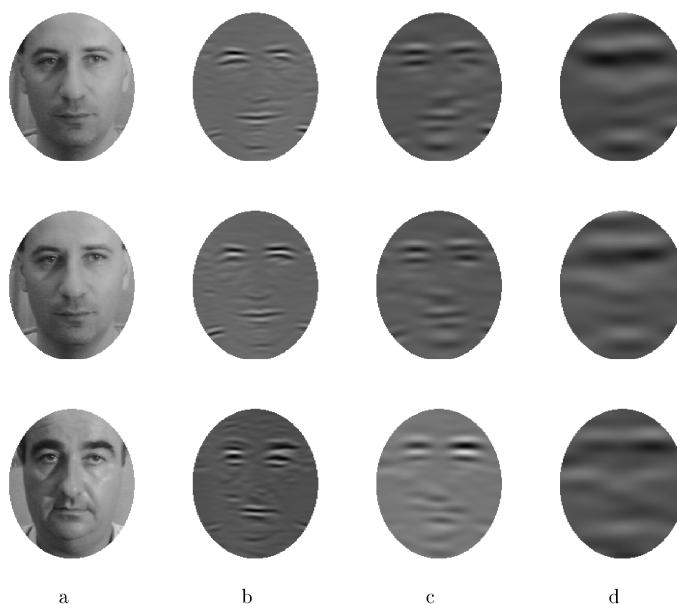


Fig. 7. Each row consists of an original face image. (Column (a)) and the image convolved with three different Gabor operators. Column (b) Horizontal Cosine Gabor with $\lambda = 8$, $\sigma = 4$. Column (c) Horizontal Cosine Gabor with $\lambda = 16$, $\sigma = 8$. Column (d) Horizontal Cosine Gabor with $\lambda = 32$, $\sigma = 16$. The upper and the middle rows are of the same face taken under left and right illumination direction, respectively. The lower row is of a different face taken under the same condition as the middle row.

from a large number of confusable faces. Fig. 7 shows three out of the 107 representations considered. The representation that compensates the best to changes in illumination in our database is shown in Fig. 7c.

We applied and compared five different distance measures to each of the image pairs. The different operator parameters affected the miss-percent by all the five distance measures in a similar manner (see Fig. 8a for the even-symmetry 2D Gabor-like filter (cosine) which is sensitive to horizontal features). However, for most of the operator parameters, the best distance measure was the local Affine-GL which measures a pointwise distance after an affine normalization of the gray-level images (see Appendix A.2). Interestingly, the simplest distance measure, the pointwise distance, was similar, in general, and, occasionally, even better, than the best distance measure. For each of the operators and distance measures, three different face masks were considered (see Fig. 4). The inner features mask was found to be the best face mask for discriminating between different faces, whereas the lower features of the face were found to be the worst (see Fig. 8b).

In our analysis of the 2D Gabor-like representation, we considered the odd and even-symmetric filters separately. A 2D Gabor representation of an image consists of a 2D wavelet transform that generates a complete set of multi-scale and multi-orientation. To study an image representation that more resembles the full 2D Gabor representation, we regarded each pixel in our representation as a vector of values. Each value corresponds to the results obtained by filtering the image with a 2D Gabor-like filter with a single scale and orientation. The distance between two corresponding pixels was considered to be the distance between the two corresponding vectors (using L_{\min} -norm). The results did not improve with respect to the single scale or

orientation that was considered throughout this paper.

In summary, for most image representations considered, the miss-percentages were above 50 percent. The representations for which the miss-percentages were below 50 percent were the ones that were sensitive to horizontal features, such as 2D Gabor-like filters, or Gaussian derivatives in the vertical direction. The lowest miss-percentage was 20 percent. Therefore, with the best operator, distance measure, and face mask that we considered, a simple recognition system will fail to recognize five out of 25 faces in our database, when the only change between a target image and the image in the database was the illumination direction. The high failure rate (about 40 percent) indicates that the failure in recognizing 20 percent of the faces in our database is significant and not the result of a small number of problematic cases.

3.2 Viewpoint

The variations between the original gray-level images of the same face that were due to a change of 34° of viewpoint direction were always larger than the variations between the images of different faces (miss-percent = 100 percent). A recognition system that is based on simple gray-level comparisons will therefore fail to recognize all the faces in our database that are due to a viewpoint change of 34° . The representations considered here were not expected to compensate well for image variations because of viewpoint. This was indeed true for viewpoint variations of 34° on the horizontal axis: the miss-percent of all the representations considered was above 50 percent.

3.3 Expression

When considering only the variation that was due to a smile in the original images, a simple gray-level comparison was found to be sufficient to recognize all the faces in

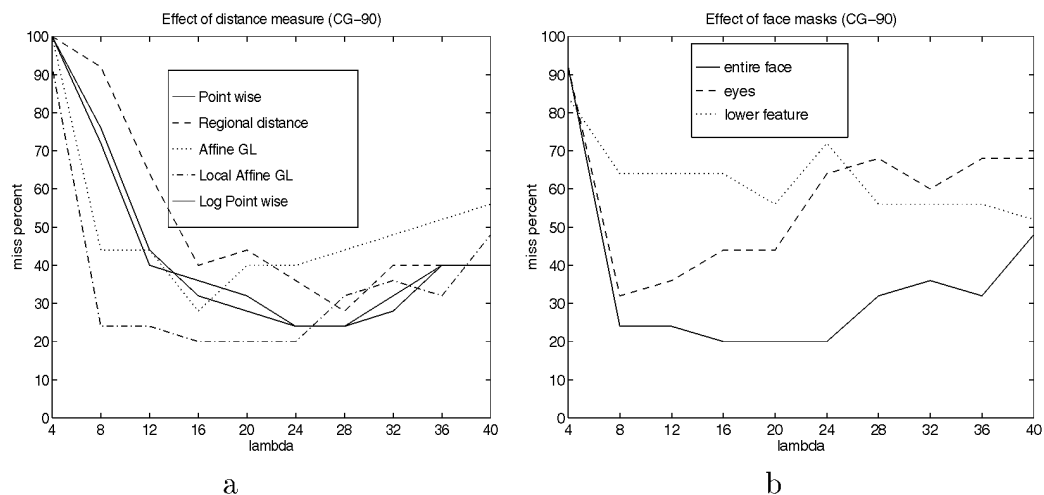


Fig. 8. (a) The effect of the five different distance measures on the miss-percent for the different operator scale for horizontal Cosine Gabor (90°). (b) The effect of different face masks (see Fig. 4) on the miss-percent as a function of the operator scale. The data is for the horizontal Cosine Gabor and local affine-GL distance measure.

our database (miss-percent of zero percent). This was true when the entire face mask was considered. However, for the other face masks (the eyes or the lower part) the miss-percentage increased to 30 percent and 60 percent, respectively. Interestingly, the recognition of the faces was impaired when several alternative image representations were considered. For example, for the image representation computed by convolving the image with the odd symmetry (sine) 2D Gabor-like function that is sensitive to horizontal features, the miss-percentage increased to 34 percent for the entire face. When variations that were due to a drastic expression were considered, the results were different. For the original images, the miss-percentage was 60 percent for the full face mask and above 80 percent for the eyes. As in the case of the smile, other image representations often increased the miss-percentage. This result demonstrates that, when a given representation is sufficient to overcome a single image variation, it may still be affected by other processing stages that control other imaging parameters such as illumination.

4 SUMMARY AND DISCUSSION

The study examined several popular image representations, computed by local operators, and often considered relatively insensitive to illumination changes applied to face images. These include images filtered with 2D Gabor-like functions, directional and nondirectional derivatives of Gaussian filters, and edge maps. The main question we addressed was to what an extent these types of image representations are sufficient by themselves to overcome image variations. To approach this problem, we constructed a database of faces in which the face identity, illumination, viewpoint, and expression were controlled separately. We considered the natural variations in the illumination, viewpoint, and expression that we face in our daily lives. The actual distances between pairs of images of a given face (or image representations) of different faces were computed and compared with the variation between pairs of images that vary because of a change of illumination, viewpoint, or

expression.

All the image representations under study were insufficient by themselves to overcome variations because of changes in illumination direction. The same result was obtained when variations between images of the same face that were due to viewpoint and expression were considered. Image representations that are sensitive to horizontal features reduced the distance between images that were due to illumination and viewpoint compared to the distance between images of different faces. This may be due to the horizontal direction of many facial features such as eyes, mouth, and eyebrows. The edges of these horizontal features are relatively insensitive to changes in illumination direction since they are due to changes in color (albedo). Furthermore, the changes in the direction of the light source and in viewpoint that were considered were both along the horizontal axis. Such changes are therefore expected to affect less horizontal features than vertical ones. In addition to the operator orientation, the scale of the operator also affects the relative distances between images of the same face relative to the distances between images of different faces. This result is supported by psychophysical evidence indicating that faces are better recognized in images with low rather than high spatial frequencies [16], [30]. However, images that resulted from convolutions with 2D Gabor-like filters with very large scales that became hardly recognizable as faces were still found to be informative for discriminating between faces in our experiments.⁴

As can be expected, when more limited variations between images of the same face are considered, the recognition increases (that is, the miss-percent decreases). It was therefore interesting to study the range of imaging pa-

4. Note that, for such scales, the external features of the face (such as the outline and background) affect the image within the face mask considered. This is because the face mask was considered after convolving the image with the appropriate operator, which is the more natural situation for natural visual systems. However, we also considered filtering the face images after extracting first the face mask. The results were that the miss-percent was above 50 percent for the best masks. This may be due to artifacts of the masks themselves (i.e., edge effects around the mask).

parameters for which the representations considered here will become sufficient for recognizing the faces in our database. As a preliminary study, we examined smaller variations between light source location, left vs. center (instead of left vs. right). The results were that for the best representations found (that were also the best in the original experiment), the miss-percent was reduced to eight percent. The failure-rate was also small, about eight percent, indicating that the misidentified faces were closer to only one or two other faces in our database. The same result was obtained when smaller variations that were due to viewpoint were considered, 17° on horizontal axis instead of 34° . Despite the improved performance, we cannot conclude that the representations studied are sufficient by themselves to compensate for illumination and viewpoint change even within these more restricted ranges. To reach such a conclusion, further studies will be required, including varying the illumination and viewpoint in directions other than the horizontal, and testing the combined effect of simultaneously changing the illumination and the viewpoint.

We next discuss the extension of the methodology presented here to more complex recognition systems. We will then present possible alternative approaches to handle image variations that are due to illumination changes. Finally, we will compare the sensitivity of the representation considered here to the sensitivity of the human visual system to changes between images of the same face.

4.1 Evaluating Recognition Systems

When developing recognition systems, it is important to analyze how well they actually cope with variations that are due to a given imaging parameter. A well-controlled database can be used to examine the system's performance. Often, when one of the representations that was considered here was used within a working recognition system, its evaluation was determined by demonstrating the performance of the entire system on a relatively small database of faces. These databases are not guaranteed to have variations between images that are due to each of the different imaging parameters. It then becomes impossible to evaluate a single component of such a system, e.g., the component that deals with variation because of illumination changes. Here we specifically examined incomplete recognition systems, however, considered the possible contribution of the early image representations schemes. The effects of illumination, viewpoint, and facial expression were studied separately. The study showed that commonly used methods for dealing with image variations that are due to illumination change are insufficient by themselves to handle relatively modest variations. In the next section, we therefore consider alternative approach to this problem.

4.2 Alternative Approaches

One possible direction for future improvement is a search for better performing representations and distance functions of the general type we have considered. However, it is also worth considering alternative approaches. We will briefly consider such approaches.

4.2.1 Independent Image Comparisons

In this approach, a face model consists of a large set of im-

ages of the same face. The recognition process consists of computing and comparing the distances between an input image and all the images comprising the model. One problem with this approach is that the number of images that the model must contain may be very large. Furthermore, when other imaging parameters, such as viewpoint and expression, are considered as well, the number of images for a given model becomes the product of the sample size for each of the spaces. Such schemes also have limited generalization capacity beyond the parameter values that were sampled and stored.

4.2.2 The Model-Based Approaches

Here the variations between different images of the same object are handled by using information specific to the object in question, such as its 3D shape and reflectance properties. The information can be stored in an explicit 3D model or, alternatively, by a number of corresponding 2D images. Such an approach is usually used to compensate for variations that are due to viewpoint changes [32], [42]. However, it can be extended to handle image variations that are due to illumination as well [15], [19], [27], [48], [49].

4.2.3 The Class-Based Approaches

Here the variations between different images of the same object are handled by using information related to a general class of objects, such as 3D shape, and the reflectance of faces in general. An example of class-based processing is extracting special facial features in a manner that is independent of the illumination direction. This can be performed by choosing, for example, only the edges that are expected to be stable in a face image from the edge map of a face. Another example is to use a special property of the class in question, such as bilateral symmetry. It has been shown [27], [36], [43] that, for bilaterally symmetric objects, certain invariance can be extracted based on a single object image. It may also be possible to incorporate knowledge about the general shape of faces to a shape-from-shading computation. In this case, the 3D shape of the object may be extracted independently of the viewpoint and illumination direction, and used for subsequent recognition.

4.3 Comparisons with the Human Visual System

Our study suggests that early image filtering and edge detection are insufficient by themselves to overcome image variations that are due to changes in the viewing conditions. It is interesting in this regard to compare the sensitivity of face identification in novel images using the methods considered here with that of the human visual system. We tested the performance of human subjects in a face identification task using the same database of faces under similar conditions [28]. Recognition was tested for both upright and inverted images. In the upright face condition, subjects were trained to recognize upright face images of several individuals and were then tested on new upright images of the same individuals. In the inverted condition, both the training and the test face images were inverted. Recognition is known to be difficult for inverted faces; however, in this study, the focus was on generalization to new viewing conditions after obtaining high recognition rates on the training set. The results showed that the identi-

fication rate for faces in upright novel images was above 97 percent correct in the range of the studied parameters. The performance of human subjects was significantly better than schemes based on the representations considered here. Processes of the type considered in this study may take part in the early processing stages in human vision, and may contribute to the object recognition process, but they are insufficient by themselves in overcoming the variations between images of the same face. It appears that, to achieve the generalization performance exhibited by the human visual system for upright face images, additional processing must take place. For inverted face images, generalization performance of human observer is reduced to 85 percent and 89 percent for new viewpoint and illumination directions, respectively. The difference between the human's capacity to compensate for the illumination changes in upright and inverted face images is consistent with the notion that compensating for illumination condition is not achieved exclusively by early processing stages such as image filtering and edge detection.

Further support for this notion is provided by brain lesion studies implying that high-level visual areas are involved in the process of compensating for illumination changes. In human lesion studies, Warrington and Taylor [44] found that patients with right posterior lesions had severe problems in identifying objects that differ in viewing positions from a training view. They also found that, for such patients, changes in illumination (in recognizing faces) were as effective as changing the viewing direction in eliciting the right-hemisphere deficit. Weiskrantz [45] made lesions to the middle and anterior inferotemporal lobe (AIT), to the prestriate, and other visual areas of the monkey. With lesions to the AIT, recognizing objects becomes very difficult even from a previously viewed image. Lesions to other parts of inferior temporal (IT) cortex and prestriate cortex did not affect the recognition of objects in images that were seen before. It did, however, affect the recognition of the same objects in novel images that varied in size or illumination. These findings suggest that the process of compensating for illumination changes requires not only the primary visual cortex, but also the participation of higher level visual areas. The results are compatible with the theory that processes such as image filtering and edge detection assumed to take place in the primary visual cortex are insufficient by themselves to compensate for viewpoint and illumination changes, and higher level processes including object and class specific processes are likely to be involved.

To conclude, overcoming image variations that are due to illumination direction is a basic problem in face recognition. Existing approaches to this problem rely primarily on universal representations, that is, representations that are not specific to faces. We examined a large number of face comparison schemes based on such representations, and we found that they are insufficient by themselves to overcome the variations that are due to illumination. We suggest that processes that utilizes more domain-specific knowledge, applicable to specific individuals or to faces in general, can be used for this purpose. The elucidation of these compensation processes will be useful for constructing better face recognition systems, and for understanding the processes used by the human visual system.

APPENDIX: METHODS

In this appendix, full details of the methods used in the experiment are given.

A.1 The Database of Faces

The database was described briefly in Section 2. Here we list more technical details regarding the images. Each image was of the size 512×352 pixels, eight bits per pixel. The image size was then decreased by half, by a subsampling. The camera ("Pulnix" TM-560 with Canon lens V6 $\times 16$ 16–100mm $F1:1.9$) was attached to a robot ("Adept One"). The camera locations were frontal, and 34° left on the horizontal axis. The distance of the face from the camera was about 110 cm. Left and right illuminations were used by turning on and off fixed light sources (see Fig. 3a and Fig. 3b). Subjects were asked to bear a neutral expression, a smile expression, and a drastic expression, and to remain still (see Fig. 3a, Fig. 3d, and Fig. 3e, respectively). The background for all images was the same: a wooden screen.⁵

A.2 Distance Measures

We considered each image representation as a gray-level image. The edge-map, which was originally a binary image, was convolved with a Gaussian (see Section A.3.1) to create a gray-level image. The following distance measures between gray-level images could therefore be used for the original images, as well as for all image representations considered here.

The following distance functions between gray-level images were considered:

Pointwise distance was defined as the average difference between the gray-level values of all pairs of corresponding pixels (i.e., two pixels in the same location). That is

$$\text{Pointwise}(I_1, I_2) = \frac{1}{n} \sum_{x \in \text{mask}} |I_1(x) - I_2(x)|$$

where n is the number of pixels in the face-mask (see Fig. 4) and $I_i(x)$ is the gray-level value of the pixel in location x in the image I_i . The face images were normalized with respect to position and size, and, therefore, gray-level values at corresponding locations with respect to the face were compared.

Regional distance was defined as the average of the minimum difference between the gray-level value of a pixel and the gray-level value of each pixel in the (5×5) neighborhood of the corresponding pixel. Formally,

$$\text{Regional}(I_1, I_2) = \frac{1}{n} \sum_{x \in \text{mask}} \min_{i \in \text{neighb}(x)} |I_1(x) - I_2(i)|$$

where $\text{neighb}(x)$ is a square of 5×5 pixels around x . Note that the regional distance compensates for a displacement of up to three pixels of the images in the plane.

5. The images were selected from a larger database, the Weizmann Face-base, that consists of 66 images for each face. The full database contains five different camera locations: -34° , -17° , 0° , 17° , and 34° on the horizontal axis; four different illuminations: left, center, right and left + right combinations; and three different expressions.

Affine-GL distance was defined as the minimum Euclidian distance between the gray-level values of one of the image and any affine transformation of the gray-level values of the other image. Formally

$$\text{Affine GL}(I_1, I_2) = \min_{a,b} \frac{1}{n} \sqrt{\sum_{x \in \text{mask}} (aI_1(x) + b - I_2(x))^2},$$

where the a and b were computed analytically. Note that the Affine-GL distance compensates for uniform affine transformation of the gray-level values of one of the images.

Local Affine-GL distance similar to the Affine-GL measure, but the minimum is computed for disjoint square blocks in the image. The size of the squares used was 16 pixels.

LOG distance The point-wise distance was computed between the \log_2 of the image representation.

A.3 Image Representations

The following image representations were considered: the original image, the edge map of the image, derivatives of the gray-level image, the image filtered with a 2D Gabor-like function, and the Log of several of the above representations. We next describe in more detail each of the operators that produced these image representations, as well as the parameters that were considered.

A.3.1 Edge Representation

Edges can be detected by maxima in the intensity gradient or zero-crossing in the second derivatives calculated by some differential operators. A number of edge detectors have been described in the literature [7], [12], [14], [20], [26], [37], [40]. We considered here the edge representation computed by the DRF edge detector [37]. (Another representation related to edges, the Laplacian of a Gaussian, is considered later, Section A.3.2.)

Unlike other representations considered here, the edge map is a binary representation. The distance functions we used were applied to gray-level images (Section A.2). Therefore, the edge maps were first convolved with a Gaussian function and then the standard distance functions were applied. Two scales of the Gaussian were considered $\sqrt{2}\sigma = 5$ and $\sqrt{2}\sigma = 11$. There are, of course, alternative methods to compute the distance between edge maps; however, we expect other methods to give qualitatively similar results.

A.3.2 Laplacian-of-Gaussian Filter

The Laplacian-of-Gaussian filter computes the second derivative of an image that was first blurred by a Gaussian function [26]. Formally

$$\nabla^2 G(r) = \frac{-1}{\pi\sigma^4} \left(1 - \frac{r^2}{2\sigma^2} \right) e^{\frac{-r^2}{2\sigma^2}}$$

The representation was computed by convolving an image with the Laplacian of a Gaussian. The scales of the Gaussian considered were $\sigma = 2$, and 4 pixels.

A.3.3 Gray-Level Derivatives

The representations of symmetrical and directional derivatives of smoothed gray-level images were considered. To compute the derivatives, the gray-level image was convolved with the derivative of a Gaussian.

Three filters were considered:

- Isotropic derivative

$$d_r = \frac{r}{\sigma^2} e^{\frac{-r^2}{2\sigma^2}}$$

where $r^2 = x^2 + y^2$.

Five values of the Gaussian standard deviation were considered: $\sigma = \{6, 8, 12, 16, 20\}$.

- Horizontal derivative (in the x direction)

$$d_x = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}$$

Five values of the Gaussian standard deviation were considered: $\sigma = \{6, 8, 12, 16, 20\}$.

- Vertical derivative (in the y direction)

$$d_y = \frac{y}{\sigma^2} e^{\frac{-y^2}{2\sigma^2}}$$

Five values of the Gaussian standard deviation were considered: $\sigma = \{6, 8, 12, 16, 20\}$.

A.3.4 2D Gabor-Like Filters

A 2D Gabor function is a product of an elliptical Gaussian times a complex exponential representing harmonic modulation [10].

Formally, a 2D Gabor function is given by

$$f(x, y) = e^{-\pi \left[(x-x_0)^2 \sigma_x^2 + (y-y_0)^2 \sigma_y^2 \right]} * e^{\frac{-2\pi}{\lambda} (\cos(\theta)(x-x_0) + \sin(\theta)(y-y_0))}$$

where x_0 and y_0 are the elliptical Gaussian center and σ_y/σ_x is the Gaussian aspect ratio. θ and λ are the orientation and wave-length of the harmonic modulation function, respectively.

We considered separately even (the real part) and odd (the imaginary part) members of the family of 2D Gabor filters with unity aspect ratio $\sigma_y/\sigma_x = 1$. These 2D Gabor-like filters have the following form

$$\begin{aligned} \text{CosG}(x, y) &= \cos \left[\left(\frac{2\pi}{\lambda} \right) (x \cos(\theta) + y \sin(\theta)) \right] e^{\frac{-(x^2+y^2)}{2\sigma^2}} \\ \text{SinG}(x, y) &= \sin \left[\left(\frac{2\pi}{\lambda} \right) (x \cos(\theta) + y \sin(\theta)) \right] e^{\frac{-(x^2+y^2)}{2\sigma^2}} \end{aligned}$$

The effects of several parameters of the 2D Gabor-like filters were studied:

- Even-symmetric vs. odd-symmetric function. That is $\text{CosG}(x, y)$ vs. $\text{SinG}(x, y)$.
- The orientation of the harmonic modulation, denoted by θ .

$\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$.

- The scale of the Gaussian (standard deviation), denoted by σ is
 $\sigma = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$ pixels, and $\lambda = 2\sigma$. The size of the mask was considered to be 2λ . The values of λ uniquely determine the scale of the operator and were used in describing the results.
- The ratio between the scale and the 2D harmonic modulation wave-length, denoted by σ/λ is
 $\lambda = \{4, 6, 8, 12\}$ for a constant scale, $\sigma = 6$, and a constant orientation, $\theta = 90^\circ$.

A.3.5 Log Representation

All of the above representations were also followed by a log function to generate additional representations.

ACKNOWLEDGMENTS

We are grateful to E. Okon and O. Smikt for technical assistance in setting up the face image acquisition system. We also thank the people who were patient enough to have long series of their snapshots taken. We thank R. Basri, N. Ferrier, D. Sagi, L. Shapiro, and J. Daugman for their comments on an earlier draft of this paper. The work was supported by the Israel Science Foundation and by the Israeli Science Ministry.

REFERENCES

- [1] C. Blakemore and F. Campbell, "On the Existence of Neuron in the Human Visual System Selectively Sensitive to the Orientation and Size of Retinal Images," *J. Physiology*, vol. 203, pp. 237-260, 1969.
- [2] B. Bruce, "Changing Faces: Visual and Non Visual Coding Processes in Face Recognition," *British J. Psychology*, vol. 73, pp. 105-116, 1982.
- [3] R. Brunelli and T. Poggio, "HyperBF Networks for Real Object Recognition," *Proc. IJCAI*, pp. 1,278-1,284, Sydney, Australia, 1991.
- [4] J. Buhmann, M. Lades, and F. Eeckman, "A Silicon Retina for Face Recognition," Technical Report No. 8596-CS, Institut für Informatik, Universität Bonn, 1993.
- [5] J. Burns, R. Weiss, and E. Riseman, "The Non-Existence of General-Case View-Variants," *Geometrical Invariance in Computer Vision*, J. Mundy and A. Zisserman, eds. MIT Press, 1992.
- [6] F. Campbell and J. Robson, "Application of Fourier Analysis to the Visibility of Gratings," *J. Physiology*, vol. 197, pp. 551-556, 1968.
- [7] J.F. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679-698, 1986.
- [8] D. Clemens and D. Jacobs, "Space and Time Bounds on Indexing 3D Models from 2D Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 10, pp. 1,007, 1,017, Oct. 1991.
- [9] J.G. Daugman, "Spatial Visual Channels in the Fourier Plane," *Vision Research*, vol. 24, no. 9, pp. 891-910, 1984.
- [10] J.G. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation, Optimized by Two Dimensional Cortical Filters," *J. Optical Soc. Am.*, vol. 2, pp. 1,160-1,169, 1985.
- [11] G.M. Davies, H. Ellis, and J.W. Shepherd, "Face Recognition Accuracy as a Function of Mode of Representation," *J. Applied Psychology*, vol. 92, pp. 507-523, 1978.
- [12] L.S. Davis, "A Survey of Edge Detection Techniques," *Computer Graphics and Image Processing*, vol. 4, pp. 248-270, 1975.
- [13] R. De-Valois and K. De-Valois, *Spatial Vision*. New York: Oxford Univ. Press, 1990.
- [14] R. Deriche, "Optimal Edge Detection Using Recursive Filtering," *Proc. ICCV*, pp. 501-504, London, 1987.
- [15] S. Edelman, D. Reisfeld, and Y. Yeshurun, "A System for Face Recognition that Learns from Examples," *Proc. European Conf. Computer Vision*, S. Sandini, ed., pp. 787-791. Springer-Verlag, 1992.
- [16] A. Fioreniti, L. Maffei, and G. Sandini, "The Role of High Spatial Frequencies in Face Perception," *Perception*, vol. 12, pp. 195-201, 1983.
- [17] V. Govindaraju, D.B. Sher, R. Srihari, and S.N. Srihari, "Locating Human Faces in Newspaper Photographs," *Proc. CVPR 89*, pp. 549-554, 1989.
- [18] N. Graham and J. Nachmias, "Detection of Grating Pattern Containing Two Spatial Frequencies: A Composition of Single-Channel and Multi-Channel Models," *Vision Research*, vol. 11, pp. 251-259, 1971.
- [19] P. Hallinan, "A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 995-999, 1994.
- [20] R.M. Haralick, "Digital Step Edges from Zero Crossing of Second Directional Derivatives," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, pp. 58-68, 1984.
- [21] B.K.P. Horn and M. Brooks, *Seeing Shape from Shading*. Cambridge, Mass.: MIT Press, 1989.
- [22] T. Kanade, *Computer Recognition of Human Faces*. Basel and Stuttgart: Birkhauser Verlag, 1977.
- [23] J.J. Koenderink and A.J. van Doorn, "Receptive Field Families," *Biological Cybernetics*, vol. 63, pp. 291-297, 1990.
- [24] B. Manjunath, C. Shekhar, R. Chellappa, and C. von der Malsburg, "A Robust Method for Detecting Image Features with Application to Face Recognition and Motion Correspondence," *ICPR B*, pp. 208-212, 1992.
- [25] S. Marcelja, "Mathematical Description of the Responses of Simple Cortical Cells," *J. Optical Soc. Am.*, vol. 70, pp. 1,297-1,300, 1980.
- [26] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proc. Royal Soc. London B*, vol. 207, pp. 197-217, 1980.
- [27] Y. Moses, "Face Recognition: Generalization to Novel Images," PhD thesis, Weizmann Inst. of Science, 1993.
- [28] Y. Moses, S. Edelman, and S. Ullman, "Generalization to Novel Images in Upright and Inverted Faces," *Perception*, vol. 25, pp. 443-461, 1996.
- [29] Y. Moses and S. Ullman, "Limitation of Non-Model-Based Recognition Schemes," *Proc. ECCV-92*, G. Sandini, ed., pp. 820-828. Springer-Verlag, 1992.
- [30] A. O'Toole, R. Millward, and J. Anderson, "A Physical System Approach to Recognition Memory for Spatially Transformed Faces," *Neural Networks*, vol. 1, pp. 179-199, 1988.
- [31] K. Patterson and A. Baddeley, "When Face Recognition Fails," *J. Experimental Psychology: Human Learning and Memory*, vol. 3, pp. 406-417, 1977.
- [32] T. Poggio and S. Edelman, "A Network that Learns to Recognize Three Dimensional Objects," *Nature*, vol. 343, pp. 263-266, 1990.
- [33] D. Pollen and S. Ronner, "Phase Relationships Between Adjacent Simple Cells in the Visual Cortex," *Science*, pp. 1,409-1,411, 1981.
- [34] D. Pollen and S. Ronner, "Visual Cortical Neurons as Localized Spatial Frequency Filters," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 13, pp. 907-916, 1983.
- [35] D. Reisfeld and Y. Yeshurun, "Robust Detection of Facial Features by Generalized Symmetry," *ICPR A*, pp. 117-120, 1992.
- [36] C.A. Rothwell, D.A. Forsyth, A. Zisserman, and J.L. Mundy, "Extracting Projective Structure from Single Perspective Views of 3D Point Sets," *Proc. ICCV-93*, pp. 573-582, 1993.
- [37] J. Shen and S. Castan, "An Optimal Linear Operator for Edge Detection," *Proc. SPIE-87*, 1987.
- [38] J. Shepherd, G. Davies, and H. Ellis, "Studies of Cue Saliency," *Perceiving and Recognizing Faces*, G.M. Davies, H. Ellis, and J. Shepherd, eds., pp. 104-131. Academic Press, 1981.
- [39] C. Stromeyer and B. Julesz, "Spatial-Frequency Masking in Vision: Critical Bands and Spreads of Masking," *J. Optical Soc. Am.*, vol. 62, pp. 1,221-1,232, 1972.
- [40] V. Torre and T. Poggio, "On Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 147-163, 1986.
- [41] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.
- [42] S. Ullman and R. Basri, "Recognition by Linear Combinations of Models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, pp. 992-1,005, 1991.
- [43] T. Vetter, T. Poggio, and H. Bulthoff, "The Importance of Symmetry and Virtual Views in Three Dimensional Object Recognition," *Current Biology*, pp. 18-23, 1994.

- [44] E. Warrington and A. Taylor, "Two Categorical Stages of Object Recognition," *Perception*, vol. 7, pp. 152-164, 1978.
- [45] L. Weiskrantz, "Visual Prototypes, Memory, and the Inferotemporal Lobe," *Vision, Memory, and the Temporal Lobe*, E.I.M. Mishkin, ed., chapter 2, pp. 13-28. New York: Elsevier, 1990.
- [46] F. Werblin, "Control of Retinal Sensitivity," *J. General Physiology*, vol. 63, pp. 62-87, 1974.
- [47] K. Wong, H. Law, and P. Tsang, "A System for Recognising Human Faces," *Proc. ICASSP*, pp. 1,638-1,642, 1989.
- [48] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *Proc. European Conf Computer Vision*, 1996
- [49] P. Viola and W.M. Wells III, "Alignment by Maximization of Mutual Information," *Proc. Fifth Int'l. Conf. Computer Vision*, 1995.

Yael Adini received her BSc in physics from Tel-Aviv University, Israel, in 1981. She received her MSc in applied mathematics and computer science from the Weizmann Institute of Science, Rehovot, Israel, in 1991. Currently, she is a PhD student in the Department of Neurobiology, Brain Research, at the Weizmann Institute of Science. Her areas of interest include lateral interactions in the visual cortex, psychophysics, and modeling.

Yael Moses received her BSc in mathematics and computer science from the Hebrew University, Jerusalem, Israel, in 1984. She received her MSc and PhD in computer science from the Weizmann Institute of Science, Rehovot, Israel, in 1986 and 1995, respectively. She was a postdoctoral fellow in the Department of Engineering at Oxford University, Oxford, U.K., in 1993-1994. Currently, she is a postdoctoral fellow at the Weizmann Institute of Science. Her areas of interests include man and machine vision.

Shimon Ullman received the BSc degree (summa cum laude) in mathematics, physics, and biology from the Hebrew University, Jerusalem, Israel, in 1973, and the PhD degree in artificial intelligence from the Massachusetts Institute of Technology in 1977. He is the Samy and Ruth Cohn Professor of Computer Science at the Weizmann Institute of Science, Rehovot, Israel. Prior to his current position, Prof. Ullman was a professor of brain and cognitive science at MIT and a member of the Artificial Laboratory there. His research interests include human vision, computer vision, and the modeling of information processing in the visual cortex. He is the author of the books *The Interpretation of Visual Motion* (MIT Press, 1979) and *High-Level Vision: Object Recognition and Visual Cognition* (MIT Press, 1996).