Int. J. Man–Machine Studies (1981) **15**, 137–178

Mechanisms of human facial recognition

ROBERT J. BARON

Department of Computer Science, The University of Iowa, Iowa City, Iowa 52242, U.S.A.

(*Received 10 March 1979*)

This paper presents an extension and refinement of the author's theory for human visual information processing, which is then applied to the problem of human facial recognition. Several fundamental processes are implicated: encoding of visual images into neural patterns, detection of simple facial features, size standardization, reduction of the neural patterns in dimensionality, and finally correlation of the resulting sequence of patterns with all visual patterns already stored in memory. In the theory presented here, this entire process is automatically "driven" by the storage system in what amounts to an hypothesis verification paradigm.

Neural networks for carrying out these processes are presented and syndromes resulting from damage to the proposed system are analyzed. A correspondence between system component and brain anatomy is suggested, with particular emphasis on the role of the primary visual cortex in this process. The correspondence is supported by structural and electrophysiological properties of the primary visual cortex and other related structures.

The logical (computational) role suggested for the primary visual cortex has several components: size standardization, size reduction, and object extraction. The result of processing by the primary visual cortex, it is suggested, is a neural encoding of the visual pattern at a size suitable for storage. (In this context, object extraction is the isolation of regions in the visual field having the same color, texture, or spatial extent.) It is shown in detail how the topology of the mapping from retina to cortex, the connections between retina, lateral geniculate bodies and primary visual cortex, and the local structure of the cortex itself may combine to encode the visual patterns. Aspects of this theory are illustrated graphically with human faces as the primary stimulus. However, the theory is not limited to facial recognition but pertains to Gestalt recognition of any class of familiar objects or scenes.

Introduction

Over the past hundred years, and particularly over the last twenty, face recognition has been the focus of numerous studies including, but not exclusively, recognition by children and adults; recognition as a function of social or ethnic class, age, sex, or beauty of appearance; recognition as a function of brain damage and cerebral dominance; recognition based on isolated facial features; recognition as a function of mode of information processing; and recognition by computer [see Baron (1979) for a comprehensive bibliography on face recognition]. Although each of these studies lends insight into the human capacity for face recognition, nowhere does there appear a description of the information processing mechanisms and networks that underlie face recognition.

The goal of this paper is to study visual perception and recognition in terms of the underlying neural networks. With this in mind, we must naturally study the brain as if it were a computer, albeit a computer of a very special kind. We must not only study the

0020-7373/81/060137+42$02.00/0

constituent networks, but also their interaction and control functions. In brief, we must study the *logical architecture* of the brain. Although this particular study will be limited to some of the networks involved with visual information processing, visual information processing networks do not function in isolation: their activity is regulated by other networks that are not directly involved with visual information. The discussion here, however, will be limited with respect to these other networks, and we will only suggest how they interact with and control the visual processing networks under consideration; we will not discuss how they work.

This study begins with a brief description of facial recognition by computer. We will analyze the essential information processing aspects of facial recognition and carefully distinguish between the information and control processes involved. We will give a brief description of a computer program for face recognition which demonstrates some of the essential processes involved, and we will summarize some computer face recognition experiments which support the theory to be discussed. After a brief discription of computer facial recognition, we will devote the remainder of the paper to a model for the networks that underlie visual perception. We will include in our analysis a discussion of various clinical syndromes of the visual system.

Instantaneous recognition of faces

When people recognize faces of friends or famous people, the recognition is almost instantaneous. In contrast, when people are presented with the task of locating one unfamiliar face among a collection of similar faces, or of determining whether an upside-down face or one presented in photographic negative is similar to a given face, or of determining whether two faces depict the same person seen from different points of view (front or profile), then processing is sequential and is based on locating and comparing features rather than processing the entire face as a Gestalt image. We will first consider the instantaneous recognition process relating to familiar faces, and defer until later a discussion of face recognition tasks that use sequential processes.

Most computer programs for face recognition are based on locating and measuring selected facial features which are then compared with corresponding measurements of known faces. Examples of the types of measurements used are ratios of distances between eye corners and mouth; ratio of height of face to width of face; and lines and angles of points along the face profile. Although these processes resemble the sequential face recognition processes used by people when performing complex face recognition tasks, they do not appear to be related to the instantaneous processes that people use when recognizing familiar faces.

One question that immediately comes to mind is: are there any computational techniques that can instantaneously distinguish between different faces? The answer is yes. Preston (1965) reported that instantaneous recognition of human faces was carried out successfully using an optical computer. Preston showed an experiment where a set of photographs of six different kings were arranged in two rows of three faces and the resulting picture was used to create a matched filter for the six faces. The matched filter is the optical computer's "memory trace", as it represents the only pictorial images that can be recognized. The filter was placed in the optical computer as shown in Fig. 1. When one of the six pictures was presented as an input to the optical computer, a bright spot was imaged in the output plane. The presence of a bright spot indicated that one of

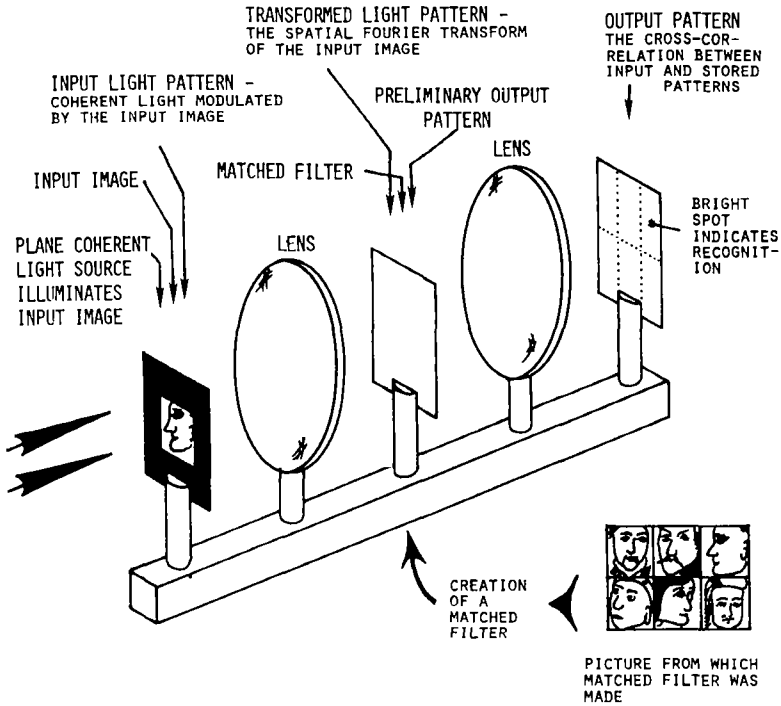


FIG. 1. An optical computer used for recognizing faces.

the six kings was presented to the computer, and the position of the spot indicated exactly which face was presented. Absence of a bright spot indicated that an unrelated picture was presented.

It is easy to show that the optical computer described by Preston computes the cross-correlation between the input picture and the picture represented by the matched filter. Cross-correlation is produced instantaneously by the optical properties of the system and is one type of template matching. The cross-correlation of two images is a measure of their spatial similarity, and for two images to be similar, they must be the same size and have the same orientation. Computing the cross-correlation function between two images amounts to comparing them against one another in every possible position, shifting both vertically and horizontally. The optical system, however, does not actually shift the images to perform the computation, and we must point out that no compensation is made for rotation and size of the input image. In order to produce a bright spot the input image must be the same size and in the same orientation as it was when the matched filter was made.

The question that may now be asked is, could cross-correlation possibly be used by the neural networks of the brain for instantaneous face recognition? Although there is currently no definitive answer to this question, one theoretical model for information storage in the brain does suggest correlation as the mechanism of association (Baron, 1970*a*), and evidence has been gathered in its support (Pribram, Nuwer & Baron, 1974.) The correlation of two neural depolarization patterns is a meaure of their spatial and temporal similarity, but without any shifting. Baron's model is consistent with

neuroanatomy as well as currently known information processing aspects of the brain, and we will show in this report that correlation has sufficient power to support face recognition as well as other visual recognition tasks.

Baron's visual information processing models

Because the models presented later in this paper are extensions and refinements of previous models, we will begin with a very brief overview of those models (refer to Fig. 2). Baron proposed that the elementary visual processing networks of the brain include

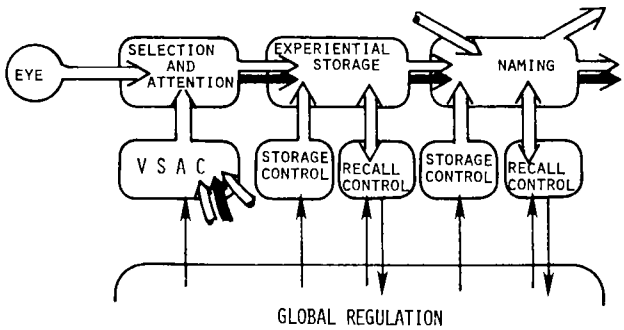


FIG. 2. The major elementary visual processing networks. The sensory processing networks process the sensory data, including visual preprocessing and selection, experiential storage, and conceptual storage (naming). The control networks regulate the activity of the sensory processing networks. The global regulatory networks regulate the gross activity of the system. Only a few of the pathways are shown.

a retina, visual selection and attention networks, an experimental memory store, and a visual naming store. The retina is an active transducer network that converts the ocular image into a *primary pictorial pattern*. The primary pictorial pattern is the depolarization pattern of the ganglion cells comprising the afferent axons in the optic nerve. The primary pictorial pattern is delivered through the optic nerve to the visual selection and attention networks for further processing. The visual selection and attention networks include the visual sensory buffers as well as networks for image enhancement, visual masking, simple geometric transformations, and visual selection. These processes and the corresponding networks will occupy the focus of much of this paper. The outputs from the visual selection and attention networks are the visual patterns that occupy the "focus of visual attention", the *secondary visual patterns* that are delivered to the memory stores of visual experience for storage and analysis, and also to the visual naming stores for elementary verbal processing (Baron, 1970*b*).

The principal control networks are the visual selection and attention control network (VSAC in Fig. 2) and a storage and recall control network for each memory store. The visual selection and attention control network determines exactly what transformations are to be performed on the primary pictorial patterns by the selection and attention networks. Each storage control network determines when and in which memory location information is to be stored, and each recall control network determines when and from which memory location information is to be recalled. The global control networks regulate the overall processing of the entire system. [For a description of some of the global control processes, see Baron (1974*a*, *b*).]

The structure and properties of the storage networks have already been described in detail (Baron, 1970*a*). In brief, information patterns to be stored arrive through a fixed input pathway, and recalled patterns are delivered through a fixed output pathway. The neurons in these two pathways are highly specified and are in a one-to-one correspondence with each other (see Figs 3 and 4). Input patterns are stored continuously in time. Each memory location can store the input pattern for some given time interval

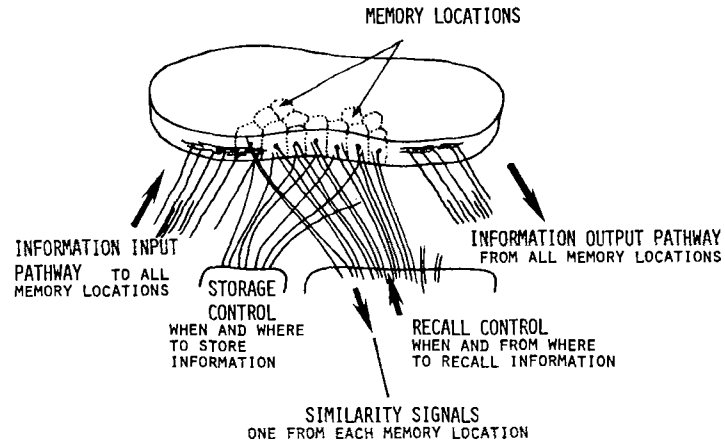


FIG. 3. The structure of an associative storage network.

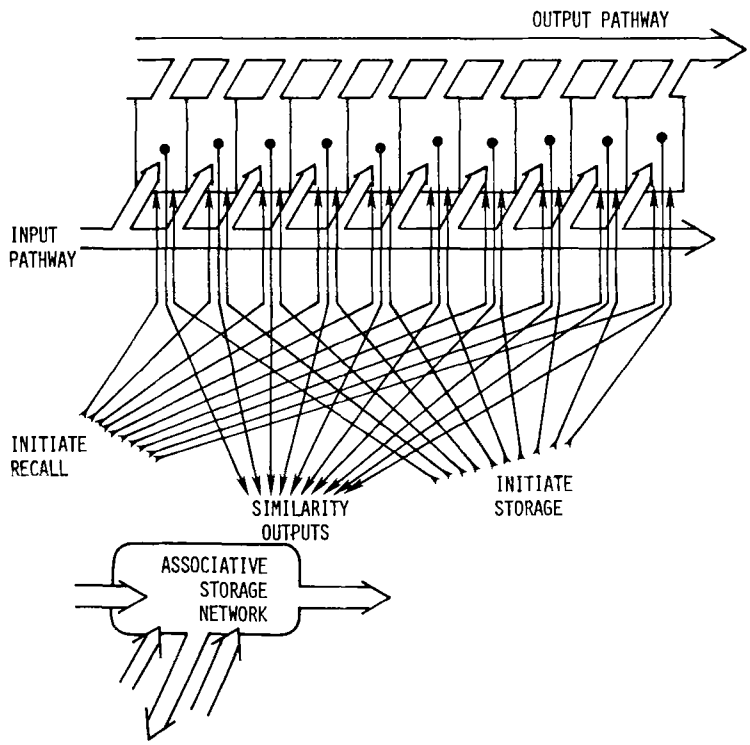


FIG. 4. The logical architecture of an associative storage network.

as determined by the storage control network and the parameters of the storage network. During recall, when a given memory location is accessed, the recalled pattern is an exact synthesis of the corresponding stored pattern. The recalled pattern varies in time exactly as the corresponding input pattern did during the interval of storage. This temporal aspect of information storage is fundamental to brain function and cannot be overemphasized.

Memory traces are encoded in the coupling parameters between certain cells in the memory stores. Storage (also called consolidation) is the modification of coupling parameters to support correlation and recall, and recall is the synthesis of an exact copy of the pattern which occurred in the input pathway during the storage interval. Parameters such as the maximum time interval that can be stored in one memory location, the number of memory locations in a storage network, and the number of cells in the input and output pathways, depend on the particular network.

The neural patterns that comprise visual experiences consist of at least two different parts: a *pictorial component*, and a *control component*. The pictorial or retinotopic component represents the geometry of the retinal stimulation. The origin of this pattern will be discussed in detail later. The pictorial component not only becomes part of the visual experience, but is also used for pictorial association and naming. The control component is an encoding of exactly how the pictorial component was derived from the primary pictorial pattern. The control component is therefore an encoding of the image origin, size, and location in the visual field. The control component is not only an integral part of the visual experience, but it is used for naming spatial relationships (to the left, above), size relationships (large, small), and surface properties (color and texture).

In order for a depolarization pattern to be recognized, it is absolutely essential that it be similar, both spatially and temporally, to a prior stored pattern. This is just as true for pictorial representations of entire faces as it is for pictorial representations of facial features such as a nose or chin. At some stage during processing, the representation of the current input pattern must be compared against the storage representations of prior input patterns, and this comparison implies some form of correlation. Furthermore, as stated earlier for brain function, correlation implies both spatial and temporal similarity measurement. It was assumed in earlier papers that the visual selection and attention networks could process the primary pictorial patterns to insure that past and present representations of visual images would be similar. In this paper we will describe how this can be done for faces.

Before describing the neural network model, we will do two things. First, we will discuss in somewhat more detail the optical computer described by Preston, and we will indicate the relationship between the optical computer and the neural network models. Second, we will briefly describe a collection of computer programs for face recognition which are based on the proposed model. These programs simulate portions of the proposed model and thereby establish computational strengths and weaknesses of the model.

The optical computer

When working with an optical computer, the following central and peripheral operations are evident.

(1) Creation of the matched filter that will be used for recognition (correlation).

(2) Creation of an input image to be recognized.

(3) Creation of a uniform coherent wavefront to illuminate the input image.

(4) Modulation of the uniform coherent wavefront by the input image to produce a wavefront whose spatial intensities are an encoding of the geometry of the input image. This is the *input light pattern.*

(5) Transformation of the input light pattern by the first lens to produce a *transformed light pattern* at the filter plane.

(6) Modulation of the transformed light pattern by the matched filter to produce a *preliminary output pattern.*

(7) Transformation of the preliminary output pattern by the second lens to produce an *output pattern* at the output plane of the optical computer.

(8) Inspection of the output pattern for a bright spot. If a bright spot of sufficient intensity is found, then its position can be used to determine what the input image represents.

Processes 4–7 are performed by the optical computer. The matched filter in the optical computer corresponds to the spatial distribution of coupling parameters in one memory location, e.g. one *memory trace,* and the computations performed by the optical computer correspond exactly to the computation performed by each memory location of a neural storage network. Unlike the optical computer which can only perform one correlation at one time, each memory location in a neural storage network correlates the input pattern against its stored patterns. It follows that a neural network performs as many simultaneous correlations as there are memory locations. The generation of a bright spot in the optical system corresponds to the rapid firing of a *recognition cell* in the storage network. A recognition cell fires rapidly if and only if the correlation between input pattern and stored pattern is high. Detection of a bright spot in the optical system corresponds to the control process which we call *recognition.* Recognition is the control process that enables recall of associated information when a pattern is presented to an associative storage network. We assume without proof that the storage, correlation, and recall processes are performed by the memory locations in the brain. Later we will present some indications that each columnar processing element in the cerebral cortex corresponds to one such memory location.

We now return to a discussion of the processes related to the creation and manipulation of the matched filter and input image for the optical computer. When using an optical computer for visual recognition, one must first create a matched filter for the images that the computer is to recognize, and one must also create the input images to be processed. The size and orientation of the images to be recognized must be the same as the size and orientation of the images used to create the matched filter, and the responsibility to insure this fact is left up to the experimenter. In addition, the experimenter creates images having uniform contrast and brightness. In fact, the same images to be recognized are generally used to create the filter.

When people recognize faces they are not so fortunate. Faces are seen at different sizes, from different points of view, under different lighting conditions, and under different contexts. It should be evident that the primary sensory patterns for a given face (the retinal patterns) are not likely to be similar from one presentation of the same face to the next, and yet the primary sensory patterns are the initial input patterns for all

visual recognition. Clearly the primary pictorial patterns must be modified before there is any hope of recognition.

We now ask, what image inhancement and selection operations are necessary in order to enable recognition? Some of of the possible operations are: compensations for image size, orientation, and lighting, and masking of the background. None of these operations are performed by the optical computer, but rather, they are performed by the experimenter as he adjusts the lighting while taking the pictures, adjusts the camera for distance, focuses it, selects film appropriate for the lighting conditions, and so forth. In the brain, corresponding operations are performed by the retina and by the networks that process the retinal image. These are the networks we have called the visual selection and attention networks, and these are the networks that will be described in detail later.

Computer recognition of human faces

When working with a computer program that is to simulate human visual processing, each brain function is simulated by a specific procedure (program) that processes one of the input images. Corresponding to the memory traces of the brain or to the matched filters of the optical computer are digital images called *templates* that are kept in the data base of the computer. The operations performed by the selection and attention networks of the brain and performed by an experimenter for the optical computer are performed by specific computer programs. The decision that we recognize an individual or that a bright spot has been formed in the optical computer is performed by a specific procedure on the computer that checks whether a correlation value is sufficiently high. The selection of specific facial features to look at, the features that "occupy the focus of attention", is an operation not performed by the optical system at all, and is performed by special computer programs that simulate the corresponding selection and attention process. Our computer programs, however, do not compensate for lighting or contrast.

In summary, the computer programs simulate each aspect of brain function, from the preprocessing of visual images to their storage in memory and finally to their correlation by the memory stores. In addition, both Gestalt and sequential recognition processes are considered for facial recognition. We must emphasize that correlation by the storage system is the only process that measures the similarity between current and past input patterns. Hence, the storage networks are solely responsible for the generation of similarity signals (familiarity). Image selection, image enhancement, size standardization, and other input processes which are essential for recognition also take place, but correlation is the only process that measures similarity and therefore directly supports recognition.

A computer system for face recognition

For our computer system, the data base consists of a collection of suitably encoded *templates* for faces and facial features which are to be recognized by the system. These templates correspond to the matched filter in the optical system and to the memory traces in the human brain. If a template for a face is not present in the data base, the face cannot be recognized. The current visual inputs are preprocessed in a way that is determined by the control system (to be described), and the resulting preprocessed

images are then correlated against all templates in the data base. If the correlation is high, further action may be taken as determined by the control system. If the correlation is very high, the control system may print a message indicating that the input face is recognized. If the correlation is not so high, the control system may try further operations to determine whether the current input pattern contains one of the faces stored in the data base. A feature extraction and comparison procedure may be initiated for this verification. In any event, the processing done on an input image is determined systematically by the control system and depends on the history of inputs to the system, the specific task the system is performing, and the correlation values for the current and past inputs. In every case, recognition is determined by correlating the processed input patterns against all stored templates. We will show that with a suitable control system and suitable processing capabilities, correlation is adequate to support the perceptual processes that underlie facial recognition.

A system for face recognition

As a first step toward a general computer vision system, the ideas described above were incorporated in a benchmark program for face recognition. All procedures were written by R. James Dawe while doing dissertation research and were coded in PL/I and executed on an IBM System 360 Model 65 computer at the University of Iowa. In the next few sections we will give a very brief overview of this system insofar as it relates to human face recognition.

THE INITIAL DATA

The data used for these studies consisted of over 150 digitized images of faces. See Figs 10, 17, and 18 for examples. These images were generated from a collection of black-and-white Polaroid snapshots of students and staff at the University of Iowa. The snapshots were full face pictures taken with the subject looking directly toward the camera. Some snapshots were taken indoors with flash while others were taken outdoors without flash. No attempt was made to control the background, what the subjects wore, hair-do's, and so forth.

The snapshots were then digitized using a Spatial Data 806 Computer Eye and TM11/TU10 magnetic tape system. The digitizing equipment was located at the University of Missouri in the Biomedical and Automation Laboratory of the Department of Engineering. Attempts were made to insure that the eyes were horizontal in the digitized image, and contrast and intensity were adjusted to be uniform. The resulting digitized images were 512 by 480 pixels (picture elements) on a grey scale of 64. These images were then reduced to 128 by 120 pixels by averaging the values in 4 by 4 squares in the original images. The resulting digitized images were used as input for the face recognition system.

With reference to the human visual system, the digitized images correspond most closely to the retinal encodings that result when looking at a face, although the retinal encoding contains much more information than ours. We assume that the human visual system compensates somewhat for intensity and contrast just as the digitization hardware did, but we imagine also that the retina performs other more sophistocated transformations on the retinal image that will be discussed later. We do not claim to be simulating the retina but only approximating its output.

LOCATING EYES IN A DIGITIZED IMAGE

In order to reduce the initial 128 by 120 image to a reasonable size for storage, the first step is to standardize the size of the face. (The issues surrounding "reasonable image size" will be discussed later.) This was done automatically by the program by locating the eyes in the input image and then reducing or enlarging the size of the input image if necessary so that the distance between the eyes was the same for all faces. Size standardization in the brain is a major part of the theory that will be presented later.

The actual procedure used for locating the eyes in the 128 by 120 input images consisted of correlating up to sixteen 20 by 23 "eye templates" against each 20 by 23 subimage of the input image (see Fig. 5). These eye templates were obtained by an

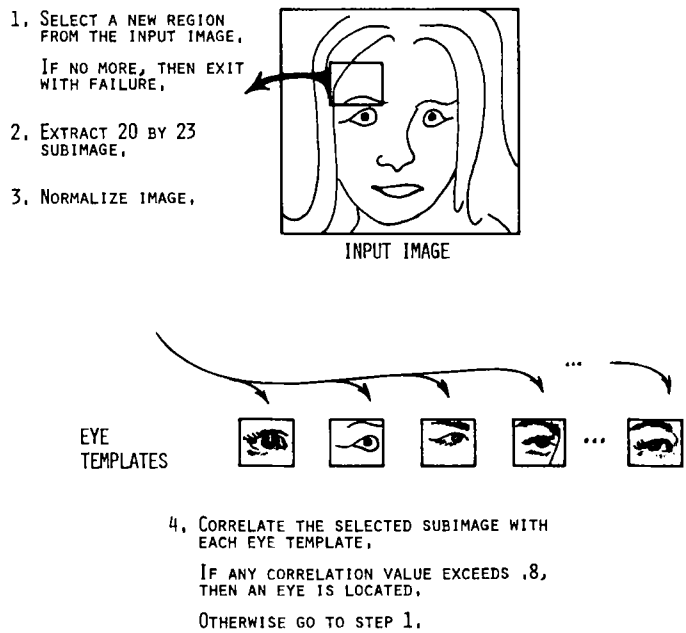


FIG. 5. The procedure for locating eyes in the input image.

adaptive procedure that incorporates the features of specific eyes into templates which match a large class of eye images. The process is similar to the feature combination process described shortly. A correlation value greater than 0·8 for any eye template indicated that an eye was located. The results compared within one row and column of the eye location determined by students when looking at prints of the digitized images.

We do not claim that this process of locating eyes corresponds closely to a brain function. We believe that the way people locate eyes is in part a reflex action based partly on the geometric shape of the eye and face. We do not know whether a template matching process like this is also used, but we suspect that one is. Our primary reason for using this technique here is that it works and is based on correlation, an essential prerequisite when simulating brain function. It also is general enough to be applied to other recognition tasks: the eye masks would simply be exchanged for masks of features that are present in the objects to be located.

THE DATA BASE

The data base has provisions for storing the memory representations of up to 75 different faces. The memory representation for each face contains up to five face features, and each face feature contains up to four distinctly different templates for that feature. Thus each face is represented by up to 20 pictorial templates. Each pictorial template in the data base has 15 by 16 pixels obtained from the input images by reducing selected areas to 15 by 16 points, and then possibly combining (to be described) the resulting image with an image already in memory.†

Associated with each pictorial feature in the data base is a control pattern that specifies the size and location of the corresponding feature in the input image. Also stored with each template is the number of input images that have been combined to create it, and also some additional information that reduces the computation time for subsequent processing (see Fig. 6).

INPUT IMAGE

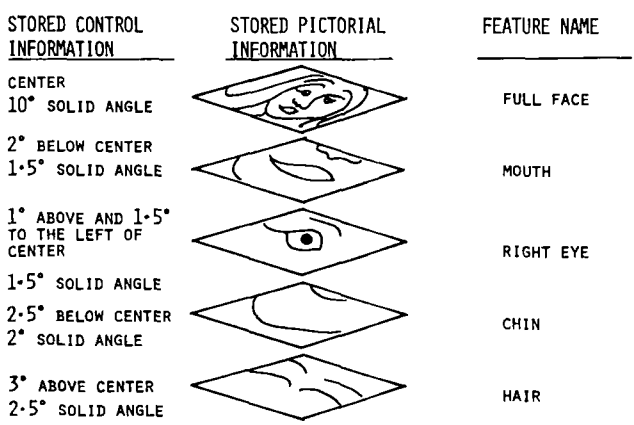


FIG. 6. The storage representation of a face.

We will later assume that a similar storage representation is used in the human brain. In particular, we assume that each human visual memory location is capable of storing a sequence of different patterns that consist of pictorial images together with their control patterns. This is analagous to the storage representation described here.

† Each storage representation requires at most 20 templates of size 15 by 16 or 4800 data points, but for a single image, only five templates or 1200 data points are used. Each input image is 128 by 120 or 15 360 data points. The storage representation therefore represents a reduction in the amount of information by a minimum factor of 12·8 to 1. As a general rule, a much greater reduction factor is evident. If six input images result in the formation of ten templates, which is typical, the reduction factor is 38·3.

The storage representation for each different face is derived from one or more input images of that person by a combination of user-supplied data and computer generated data. For each face, the user supplies a list of features of interest. Each feature is specified by giving its co-ordinates and size in the input image, and possibly its name (nose, chin, etc.) for bookkeeping purposes. The co-ordinates of a feature are the row and column numbers of its upper-left corner in the input image, and its size is the number of rows and columns it takes in the input image. The selected area containing the feature is reduced to the standard 15 by 16 size for storage.

We must emphasize that a "facial feature" is simply an image derived by selecting a particular area in the input image for further processing. There is no other significance attached to the concept. A face image can lead to a Gestalt or immediate recognition of a face in just the same way that a face feature image can lead to a Gestalt recognition of a feature.

The only process that was not automatic was the selection of features to be used during sequential recognition. The experimenter took the digitized images and selected features which to him were interesting, such as the eyes, nose, chin and so forth, and these features were then used during subsequent processing. Personal experience certainly plays an important role in a human's choice of features, but there is no corresponding automatic mechanism for selection built into our computer programs.

When the first picture of a person is processed, the computer locates the eyes and standardizes the size of the picture as described earlier. The entire face image is then reduced to 15 by 16 pixels for storage, and this 15 by 16 *full face image* or *Gestalt image* becomes the first feature image in the face representation. We note that the full face image can lead to immediate facial recognition in exact correspondence to the optical system described earlier, and also in correspondence with immediate human recognition. The remaining features specified by the user are then obtained and stored, and the appropriate control information is automatically added to the storage representation.

If a second image of the same person is incorporated in the data base, processing is somewhat different. After the face is standardized, the full face image is reduced to 15 by 16 and correlated with the stored full face representation. If the correlation value is above a constant called the *recognition threshold* the data base is not modified at all. (One might say that if a face is already sufficiently familiar, the program pays no attention to it.) If the correlation is between a constant called the *combine threshold* and the recognition threshold, the current 15 by 16 image is combined with the corresponding stored image. (In this case, the novelty of the image is high, so that the program does pay attention to it.) The combination is obtained by replacing each stored pixel in the template by the weighted average of the stored pixel value and the new pixel value. The weight is determined by the number of images that have already been combined, and that number is then incremented.† If, however, the correlation is below the combine threshold, the entire image is added to the file as an alternate full face image. This leads to an alternate Gestalt image for the same person.

† The process of template combination corresponds to the modification (adaptation) of memory traces. If there is an analagous network in the human brain, then this analysis suggests that it should have adaptive memory traces. Since it is our belief that the memory traces of experience are not adaptive, this suggests the possibility that the storage system that holds the storage representations of faces is different from the experiential storage system.

After the full face image is processed, the remaining features are processed in the same way. However, the features are now determined from the control vectors in the stored representation rather than from user supplied data. That is, the incorporation of subsequent images of a person is fully automatic.†

When the data base was completed, it represented 42 different people and was created by processing 89 different 128 by 120 pixel input images. The recognition threshold was set at 0·8 and the combine threshold was set at 0·7 when the data base was created.

FACE RECOGNITION

An input image is *recognized* by the face recognition system if it is correctly identified as one of the people in the data base. An input image is *mistakenly recognized* if it is incorrectly identified as on of the people represented by the data base, and a face is *missed* if it is an image of a person in the data base, but it is not identified as such.

In order to identify an input image, the eyes are located and the face standardized as described earlier. The full face image is then extracted and reduced to 15 by 16 for comparison with the data base. The resulting 15 by 16 full face image is correlated with every full face image in the data base. This corresponds exactly, we believe, to the first step in human face recognition. Corresponding to each storage representation are two values, a high value called the *threshold of recognition*, and a lower value called the *threshold of recall.* If the correlation value between the input pattern and the full face template in one of the storage representations exceeds its threshold of recognition, the system responds immediately with a statement of *recognition*, and the name associated with the responding storage representation is given as the name of the input. Otherwise each storage representation for which the correlation value exceeds the threshold of recall is said to *respond* to the input pattern, and the set of all such responding representations is the *conflict set.*‡ If the recognition does not occur, the system attempts to verify that the input image corresponds to one of the images in the conflict set. That representation in the conflict set with the highest correlation value is checked first.

Verification is the process of determining if the input face is the same face that is in the storage representation, and proceeds as follows. The stored control information for each feature in the storage representation is recalled and used to select the same feature from the input image. That feature is correlated against the corresponding stored feature template. If for at least three out of four of the features one of the resulting correlation values exceeds the threshold of recall, the face is identified and the process terminated. Otherwise, verification is attempted for the next face in the conflict set.

There is a direct correspondence between recognition by this system and recognition in Baron's model. Here, each representation in the conflict set is used to guide the search of the input image, and during the search, correlations are performed only on the

† In the computer programs, the representations consist of a linked collection of images. Since we assume for human storage that each memory location holds a sequences of images but not a structure, the corresponding neural representation for a face would consist of a set of sequences of images, one sequence in each memory location.

‡ The term "conflict set" is borrowed from production system terminology because of the close correspondence between these two systems (Newell, 1973). In this model, the full face image is the entry condition and the verification procedure described shortly is the corresponding action.

images stored in that representation. In Baron's model, the processing of sequences of images is a consequence of the structure of the storage network. Furthermore, incoming images are correlated against all corresponding images throughout the storage network. Thus the process is much faster. Since it is assumed for the model that incoming images may be stored in memory at the same time or immediately following presentation, a face that was not seen before could have a storage representation created immediately following an unsuccessful verification. As a result, many faces would share the same sequence of control patterns, and during verification, many faces would be verified simultaneously.

COMPUTER FACE RECOGNITION EXPERIMENTS

Experiments were conducted to determine the performance of this system. These experiments include the following:

effect of size of template on recognition accuracy;
effect of changing the recognition and combine thresholds on the storage representation;
effect of changing the lighting on recognition and storage representation;
effect of face rotation on recognition;
effect of changing the size of an input image on recognition.

Among the notable results were the following: a recognition accuracy of 100% was achieved for a data base consisting of 42 faces which was created with a recognition threshold of 0·8 and combine threshold of 0·7. Over 150 different faces were then presented to the system, including the faces used to create the data base. All of the original faces were recognized, and faces not in the data base were rejected. The recognition accuracy was not enhanced by using templates larger than 15 by 16. In fact, the 15 by 16 size seemed to contain the essential Gestalt information for recognition, and increasing the size substantially increased the execution time with no particular advantage. Changing the values of the recognition and combine thresholds changed the storage representation as follows. When the combine threshold is decreased, fewer new images are created since more images are combined into a single template, but this resulted in lower correlation values when the same input images were later used as an input. The system therefore becomes less discriminatory. When the recognition threshold is decreased, fewer images are incorporated in the templates. The result is that different pictures of the same face result in lower correlation values and are missed during recognition. Different lighting conditions have a notable effect on the size of the storage representation for a given person. Pictures taken with flash tend to be recognized by templates that were made from other pictures taken with flash whereas pictures taken without flash tended to correlate below the combine threshold and therefore resulted in new storage representations. Said another way, if the data base is created only with pictures taken with flash, then it does not recognize the same faces when taken without flash, and vice versa. Face rotations of up to 20 degrees did not affect recognition while rotations beyond 20 degrees were not recognized. By creating a new face representation with the face viewed at a 20 degree rotation, the system could then recognize all rotations up to 35 degrees. Finally, large changes in the size of the input image could not be adequately processed.

SIGNIFICANCE OF THE COMPUTER SIMULATION RESULTS TO HUMAN FACE RECOGNITION

We have described various information processing operations that relate directly to human face recognition. These include location of the face within the visual field, determination of its size, standardization of its size, correlation of the standardized input image with stored representations of known faces, location of facial features based on information contained in the storage representations of known faces, correlation of features of new faces with corresponding features of stored representations, and threshold analysis of the correlation values for determining whether a recognition criterion is satisfied. Each of these operations is performed in the brain and we will show in the remainder of this paper how the operations can be performed in a natural way in neural networks.

The first conclusion to be drawn from the previous discussion is that the operations described are sufficient to account for most of our ability to recognize faces. The second and more significant conclusion is that the number of cells involved in the brain's storage representation, its *dimensionality*, need not be very large and probably is not. We showed that storage representations having templates whose sizes were 15 by 16 (240 image points) are sufficient for recognition based on correlation, and in fact, earlier experiments indicated that images having larger dimensionality were no better and often not as good. This agrees with estimates on the size of neural patterns in permanent memory based on the architecture and connectivity of the cerebral cortex (Pribram *et al.*, 1974). This also agrees with studies by Harmon (1971, 1973) showing how little information is required for people to recognize pictures of faces. In brief we are justified in assuming that the human storage representation of the pictorial component of permanently stored visual patterns consists of no more than a few hundred elements. The detail in memory derives from having many small images together with the appropriate control information and not in having large detailed representations of the visual world.

Human networks for face recognition

Our fundamental assumptions regarding early (automatic) stages in human visual information processing during face recognition are these.

(1) Recognition results from the comparison of current neural depolarization patterns with stored representations of prior neural depolarization patterns. Storage networks perform the comparisons and are the only source of similarity information in the brain.
(2) The storage representation of permanently stored visual information has two components, a pictorial or retinotopic component, and a control component. The pictorial component is an encoding of the geometric properties of the primary pictorial pattern, while the control component indicates exactly how the pictorial component was derived from the primary sensory pattern.
(3) Primary visual patterns (retinal patterns) are processed by the visual selection and attention networks to produce the pictorial component of the secondary visual patterns which are then sent to the storage systems for storage and analysis. Analysis is the comparison for similarity of current input patterns with stored representations of prior input patterns.

(4) In order for a storage network to produce a strong similarity signal based on the retinal patterns, the current secondary pictorial pattern must be similar, both spatially and temporally, to a prior stored secondary pictorial pattern.
(5) The control component of the storage representation of visual experience is derived from the control patterns which determine exactly how the visual selection and attention networks are to process the primary visual patterns.

Our goal for the remainder of this paper is to study the various neural networks that support the processes described above. We will suggest specific network architectures and visual operations, but we will not as a rule indicate in which specific networks of the brain the operations take place. As a consequence of the proposed architecture, we will show how damage to various processing units can result in well known clinical syndromes of the visual system.

Before suggesting specific network architectures, we must caution that many of the processes are performed independently by several brain systems. One of the largest obstacles to overcome when attempting to understand the brain as a computer is the masking of function due to interactions between various independent subsystems that share in the responsibility for each task. Within the visual system there is co-operation and interaction between many different neural subsystems at many different levels of processing. Within the retina are at least two subsystems: the cone–h/bipolar–s/ganglion cell system, and the rod–d/, e/, or f/bipolar–s/ganglion cell system (Krieg, 1966). Within the primary cortex are the ocular dominance system, which appears to relate to three-dimensional processing, and the orientation-specific cell systems (Hubel & Wiesel, 1977). Eye movements are under the control of the saccadic system, the smooth pursuit system, the vergence system, and the vestibular system (Dick, 1976). At least two modes of visual perception suggest different underlying neural mechanisms: the proximal mode mediated by visual–pictorial (retino-topic) stimulation, and the visual constance mode mediated either by object–object relationships (an object moving in space), or by subject–object relationships (Mack, 1978; Lee, 1978). Several different visual memory subsystems are also evident: immediate memory (visual sensory buffers), temporary or short-term memory, and permanent or long-term memory. Analysis of visual information has many different components: visual–pictorial (shape and form), color, texture, size, and spatial location, to name a few, and each of these capabilities is mediated by distinctly different neural subsystems.

At a somewhat different level, visual processing appears to be regulated by several different subsystems characterized either as automatic or controlled (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Recognition may be Gestalt (immediate), which is generally thought to be right-hemispherical, or it may be sequential (feature by feature), generally thought to be left-hemispherical. Moreover, different search strategies appear to be learned and used by different individuals during visual search (Carey, 1978; Carey & Diamond, 1977; Corcoran & Jackson, 1977; Gould, 1973; Hailine, 1978; Howe, 1970; Just & Carpenter, 1976; Maurer & Salapatek, 1976; Neisser, 1964; Prinz, 1977; Walker-Smith, Gale & Findlay, 1977). Visual selection may be mediated either by eye movements or by an internal attention process (Sperling & Melchner, 1978), and the choice of what occupies the focus of attention depends both on the mode of processing and on the task being carried out.

We now ask the following questions:

(1) What processes must be performed by the visual selection and attention networks in order to enable face recognition?
(2) what kinds of neural networks can perform these processes?
(3) what are the corresponding networks of the brain?
(4) how are these networks controlled?
(5) what control strategies are used, and what processes are involved in the selection of control strategy?

WHAT PROCESSES UNDERLIE FACE RECOGNITION?

Based on our earlier discussion, certain processes seem to be implicated in facial recognition.

(a) Encoding of the ocular image into patterns of depolarization that represent the visual field.
(b) Selection of specific features in the visual field upon which to focus the attention. In the case of face recognition, the eyes are an obvious candidate for an initial choice.
(c) Movement of the eyes, head, and body so that the selected features occupy the focus of visual attention, even if momentarily.
(d) Possible preprocessing of the retinal pattern to compensate for different brightnesses, contrast, and shadows.
(e) A possible masking of the background.
(f) Delivery of the processed visual pattern to the storage networks for storage and analysis.
(g) Analysis of the correlation signals and determination of immediate recognition.

The above processes may be considered "automatic" since they do not depend on detailed knowledge of what is in the visual field. Moreover, if the subject is performing a face recognition task, there may be a predisposition to locate facial features within the visual field. This is because the subject is able to use knowledge about faces from memory. This aspect of facial recognition is beyond the scope of this paper. (See, for example, Mackworth & Morandi, 1967.)

Once the subject has located a face in the visual field and the facial pattern has been extracted and standardized in size, recognition by the storage networks is possible. If the face is familiar (known to the subject), then the correlation values from memory should be high. High correlation signals not only result in the sensation of familarity, but under normal circumstances they enable recall of information about that person—why he or she is familiar, where he or she was last seen, and so forth. The process of naming the individual, however, differs since it requires access to and recall of verbal rather than experiential information.

SEQUENTIAL PROCESSES IN FACIAL RECOGNITION

Once a face has been isolated in the visual field by the instantaneous processes detailed above, different operations are involved in subsequent processing. If the person is

familiar and is immediately recognized, then the subject may have access to associated experiential information about the person. We say that a person (face) is *recognized* when associated information is made available for recall. The question that must now be asked is, what happens when the initial full face (Gestalt) image does not result in the ability to recall associated information? That is, what happens if the person appears familiar but is not recognized? Evidence cited earlier suggests that the subject enters into a sequential processing mode, looking at various facial features, in order to determine who the unknown person is (see, for example, Walker-Smith *et al.*, 1977). This is analagous to the sequential process described earlier in our computer programs.

During sequential processing, knowledge about the subject being analyzed (whether it is a face, place, object, etc.) plays a central role in determining how subsequent images in the visual field will be selected. Said another way, the focus of attention (including but not exclusively the scan path) is guided by available knowledge about the nature of the object in the visual field. This implies that the control patterns which determine the focus of attention are stored in a permanent memory system, and these stored patterns are recalled and used to guide the search of the visual field (see Noton, 1969; Noton & Stark, 1971*a*, *b*).

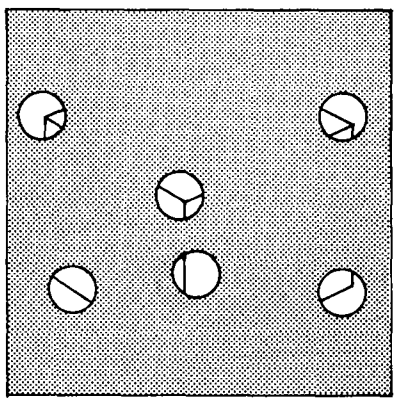


FIG. 7. An object hidden behind a piece of cardboard.

As a demonstration, consider the object hidden behind the cardboard in Fig. 7. The cardboard, shown as a speckled region, has several holes in it. In this example, if either of the lower-left holes was selected first, the information is so uninformative that one of the other holes would immediately be selected.† If the upper-left or central hole was selected, then a hypothesis about the identity of the hidden object may be formed and the verification mode of processing entered. For this figure, it is most logical to guess that the hidden object is (an isometric drawing of) a cube. Appearance of a "corner feature" is generally sufficient to enable recall of the storage representations of images of cubes. The corner feature in this example plays exactly the same role that the full face

† Some initial feature must be recognized before a guess can be made as to the identity of the object. Until a feature is recognized, the visual field can only be searched in a way that is independent of specific objects in the visual field. There is good evidence that such searches are not random but are either based on knowledge about the object being sought or the reason for the search. The control processes that underlie this type a visual field search are beyond the scope of this article.

image does when recognizing faces—it enables access to associated information. Once access to storage is enabled, the stored information can be used further to direct the search of the visual field. In this example the goal becomes that of verifying whether the hidden object is a cube or not.

MEMORY VERIFICATION

During *memory verification*, the patterns stored in one of the associated memory locations (i.e. the conflict set) are recalled, and the control components of the recalled patterns are used to guide the search of the visual field. In particular, the stored control components indicate exactly how the visual field was originally scanned. As such, they are an encoding of the sequence of selected features, their size, and how they were extracted from the visual field. If these control patterns are recalled and systematically modified, the current visual field can be searched in the same way. What we mean by systematic modification is this: if the object represented in memory was farther away from the subject, then corresponding features would have appeared closer together and smaller in the retinal image; if the selected regions in the current visual field are systematically chosen farther apart, and the corresponding retinal images reduced in size by the correct amount, the same features would occupy the focus of attention. The secondary visual patterns would then be spatially and temporally similar to the patterns stored in memory, and high correlation signals would result for the entire input presentation. This sequence of high correlation signals can then be used by the control networks to conclude that the current and past objects are the same.

Referring to the hidden cube described earlier, Figs 8(a) and (b) show two different situations where verification would not be satisfied. In Fig. 8(a), all the features except the central one are similar, but the central feature differs from what is expected based on a "cube" hypothesis. In Fig. 8(b), the features are all identical, but they are in the wrong places and hence the order in which they would be scanned is wrong.

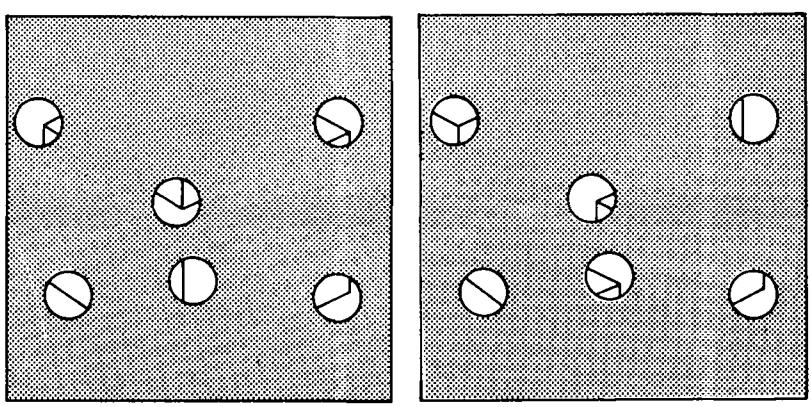


FIG. 8. Two cases where a "cube" hypothesis fails.

We note here that it is absolutely essential for memory verification to be able to determine the size of the object in the visual field. In order to determine its size, it is necessary to determine the distance between two features of the object—the distance between the eyes, for example, or the size of the head. (Knowing the size of a head is

equivalent to knowing the height—e.g. the distance from top to bottom.) This implies either that the entire object is recognized in the Gestalt sense (i.e. it is a face), or that specific features are recognized in the Gestalt sense (e.g. the eyes, the corners of a cube). We suspect that several different mechanisms are used depending on the size (distance) of the object from the subject, the nature of the object, the context of the situation, and so forth.

It should also be clear from this discussion that the sequences of events that underlie facial recognition, except in the very simplest cases, are very complex. Furthermore the sequences of events (recall of permanently stored visual control patterns, selection of the primary visual field patterns to be analyzed, and so forth), must have been learned at some time in the past. *Global control sequences* of this type are evident in all intelligent processes and must themselves be stored in some memory system in the brain. [A brief discussion of this topic is given in Baron (1974*a*,*b*), but also see Carey & Diamond (1977) and Maurer & Salapatek (1976). Also see Senders, Fisher & Monty (1978).]

NETWORKS FOR SIZE STANDARDIZATION

In the next few sections we discuss various techniques and neural architectures for processing the retinal patterns, including standardizing the size of an input pattern and selecting objects for the focus of attention. We suggest in part a specific correspondence between the proposed networks and those of the human brain

SIZE STANDARDIZATION

There are several different computational techniques that can be used for reducing the size of a neural pattern without changing its geometric (pictorial, topographic) properties. If the input pattern does not differ substantially in size from the desired output pattern, then a network of the type illustrated in Fig. 9 can be used. Figure 9(a) shows the cross section of a network having a single collection of input neurons and a single collection of output neurons. The axon ramifications of a typical input neuron are indicated as a cross-hatched region, while the dendritic ramifications of a typical output neuron are indicated as a speckled region. Potentially, each input neuron can be connected to each output neuron. We assume, however, that unless a regulatory neuron as described below is firing, the input neurons do not influence the firing rates of the output neurons.

Figure 9(b) shows an additional *regulatory neuron* whose function is to enable the input neurons to depolarize the output neurons. The regulatory neuron is cross-hatched in the figure. When the regulatory neuron fires, the coupling between input and output neurons whose axon and dendrite ramifications intersect the axon ramifications of the regulatory neuron are effectively connected, so that when an input neuron fires, the corresponding output neuron fires. The regions of coupling are speckled in the figure. As is evident in Fig. 9(b), when the single regulatory neuron shown fires, each input neuron effectively excites a single output neuron. The output pattern is simply a copy of the central part of the input pattern.†

† Regulation may work in exactly the opposite way. Regulatory neurons may inhibit the local coupling between input and output cells, and coupling only occurs when a regulatory neuron stops firing. This distinction is not important to this discussion. [See Beaudet & Descarries (1978) for a discussion of possible neural mechanisms.]

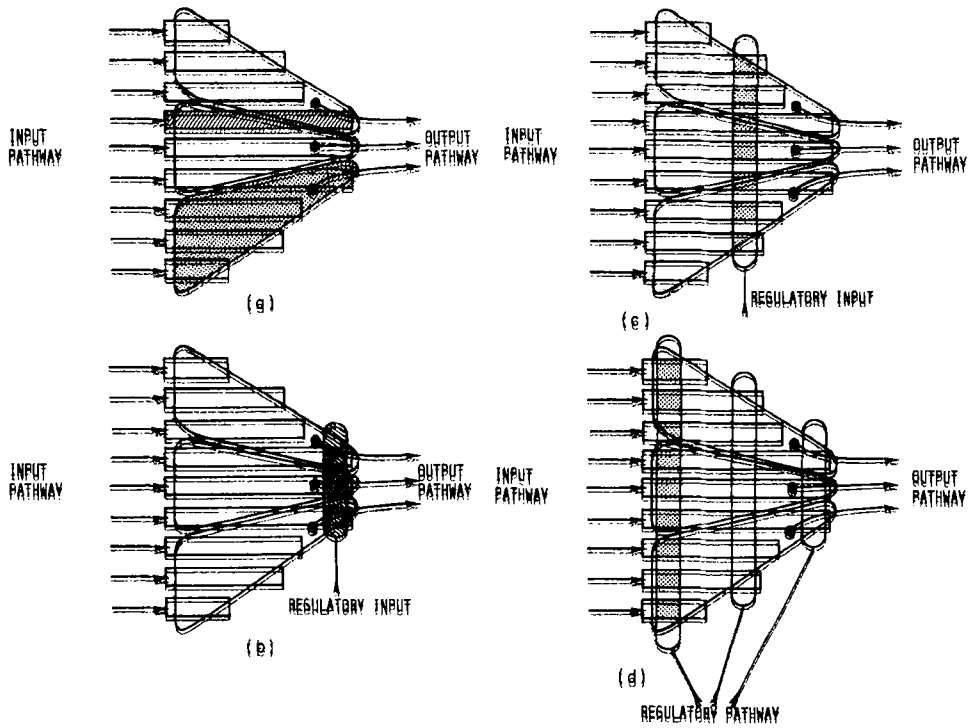


FIG. 9. (a) The input (cross-hatched) and output (speckled) cells for a size reduction network. (b) One regulatory neuron (cross-hatched) is shown, and the regions of coupling (speckled) are indicated. (c) A regulatory cell that reduces the size of the input pattern by about 2/3. (d) An entire reduction network.

Now look at Fig. 9(c). When the single regulatory neuron fires, the input neurons and output neurons are connected in a different way from that in Fig. 9(b). The regions of coupling are speckled in the figure. In fact, now two or three input neurons excite a single output neuron so that the firing rate of an output neuron is proportional to the sum of firing rates of the corresponding three input neurons. For the network in Fig. 9(b), the output pattern represents only a small part of the input pattern but with the same detail, while for the network of Fig. 9(c) the output pattern corresponds to three times the linear size of the input pattern but with correspondingly less detail. The input pattern is reduced. Figure 9(d) shows a third regulatory neuron that reduces the detail of the input pattern even further but increases the extent of the input pattern represented. Figure 10 illustrates the processes of size reduction and selection, and also shows a schematic diagram of networks for size reduction and selection.

A second network for size standardization will now be described that is based on entirely different principles. Recent research by Schwartz and colleagues (Schwartz, 1977*a*,*b*; Weiman & Chaikin, 1977) suggests that the connections between the retina and the primary visual cortex are arranged in such a way that the retinal topology is transformed according to a complex logarithmic mapping. One interesting fact about the transformation is that a change in size of the retinal pattern about the center of the transform results in a shifted cortical pattern, but one whose size is not changed. Also, a rotation about the center of the retinal pattern also results in a shifted cortical pattern,

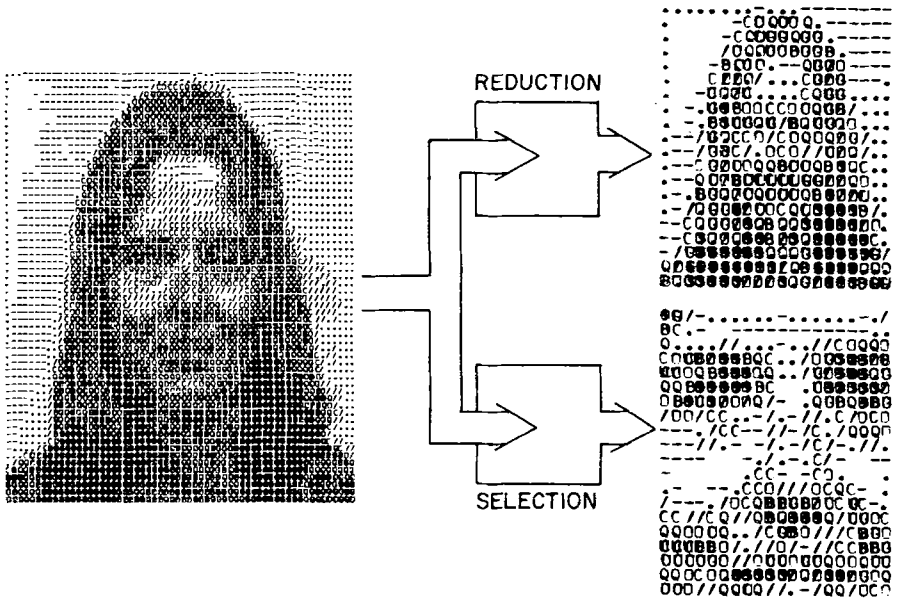


FIG. 10. An input image and two output images, one reduced in size by 2/3, and the second showing a region selected from the center of the input image. The input image is 60 by 60 and the output images are both 20 by 20.

but this time shifted in a direction perpendicular to the previous shift. A second fact of importance is that the central portion of the visual field in the transformed visual field pattern occupies a much larger percentage of the pattern. This means that the central part of the visual field will play a much stronger role in recognition (association) than the surrounding portions do. Figures 11–14 illustrate some of these relationships.

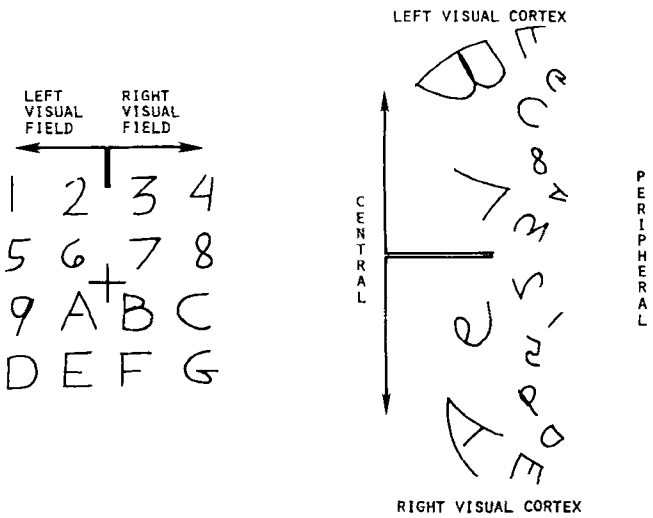


FIG. 11. An input image and the appearance of the same pattern at the primary visual cortex (according to Schwartz, 1977*a*,*b*). The center of fixation of the input field is indicated by a "+". If the same pattern were increased or decreased in size, the cortical pattern would be shifted to the left or right. If the same pattern were rotated about the center of fixation, the cortical pattern would be shifted vertically.

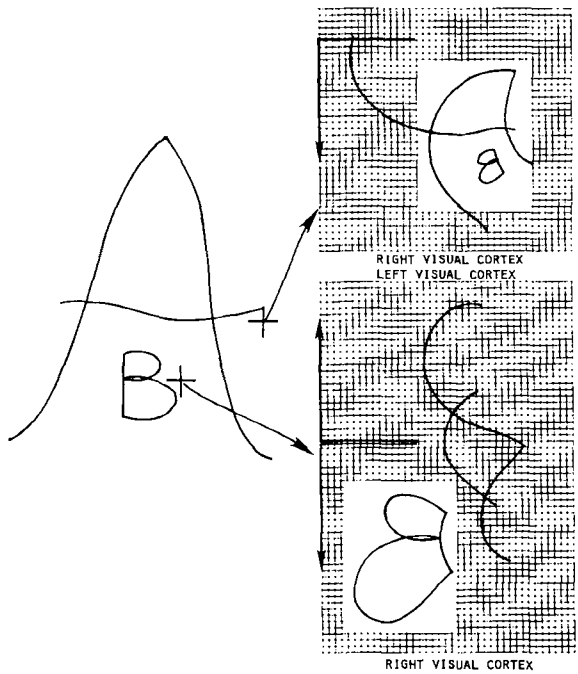


FIG. 12. An input image and the cortical representations for two different fixation points. Regions of the cortical representation selected for the focus of attention are indicated (compare with Fig. 13).

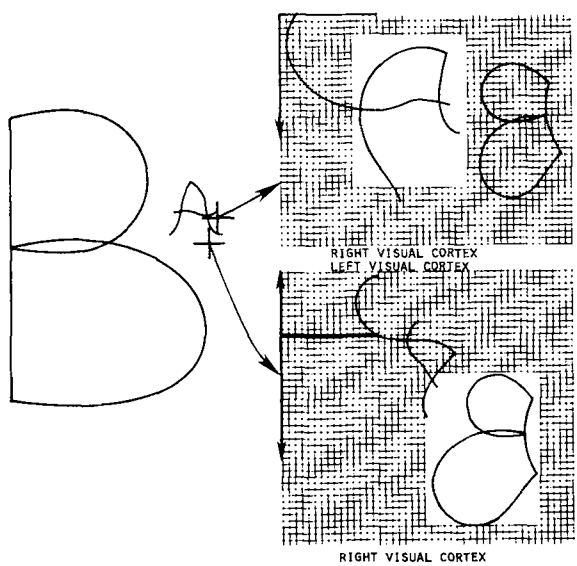
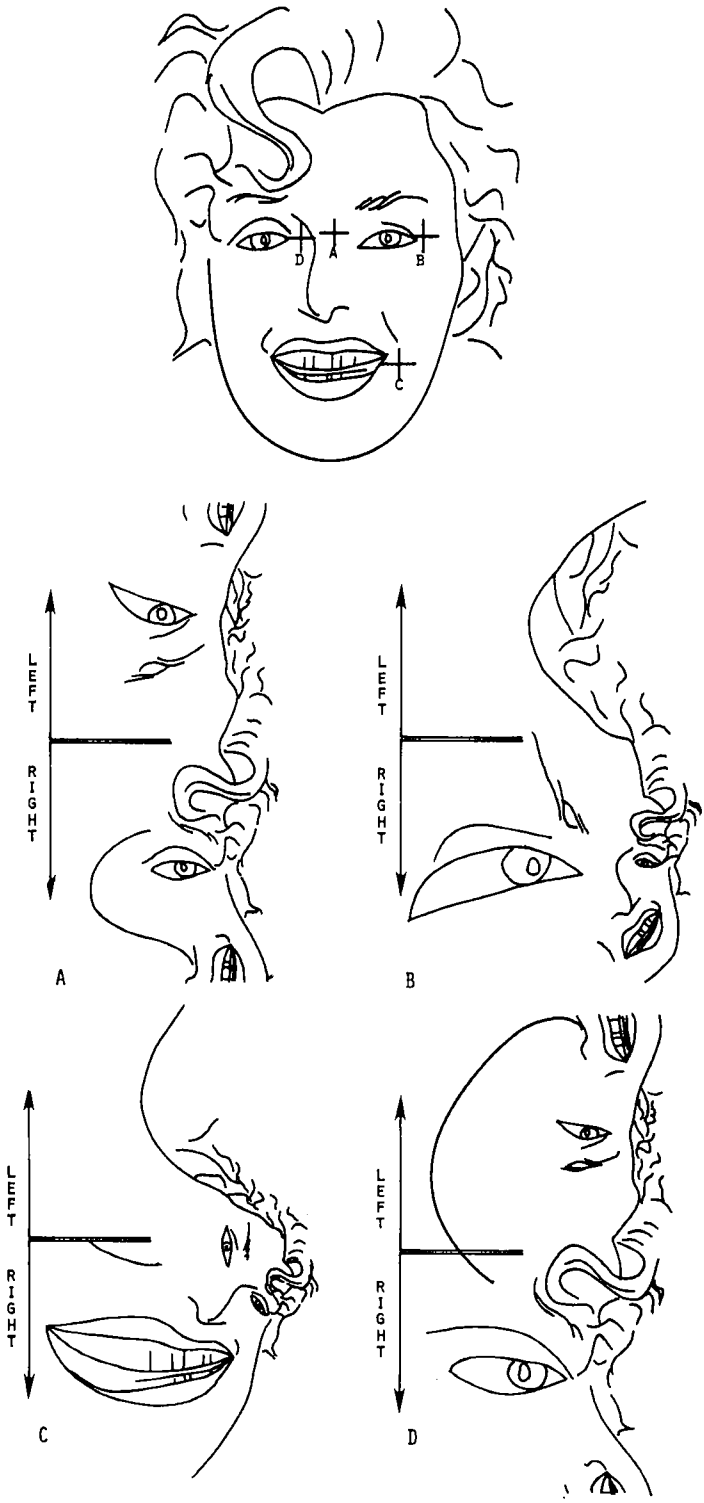


FIG. 13. A second input image and the cortical representations for two different fixation points (compare with Fig. 12).


D
A
B
C
LEFT
RIGHT
A
LEFT
RIGHT
B
LEFT
RIGHT
C
LEFT
RIGHT
D

For size changes or rotations of the retinal image, the retina–cortex connections simply produce a shifted cortical pattern. As a consequence, in order to produce a secondary pictorial pattern that is independent of size or rotation of the retinal image, it is only necesssary to *select* a corresponding pattern from the cortical cells. The problem of limited size reduction or rotation is therefore changed into the much simpler problem of selection. Figures 12 and 13 illustrate the selection process.

Figure 15 shows the cross section of one type of selection network. In Fig. 15, if the first control neuron fires, the left-most portion of the input pattern is selected; if

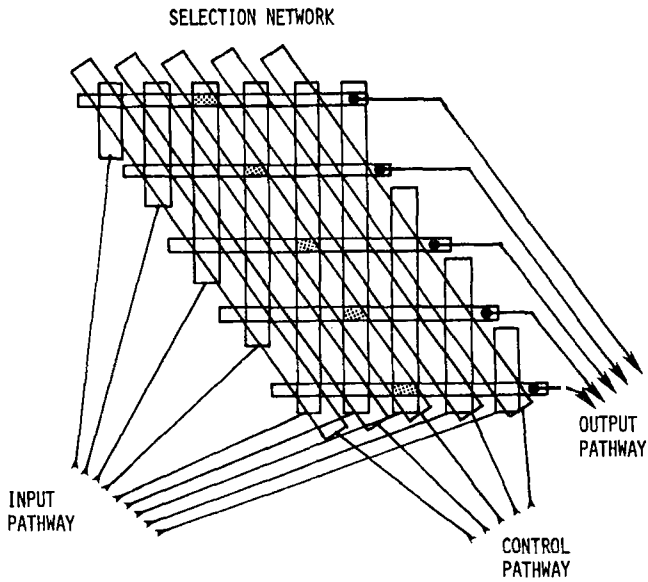


FIG. 15. The logical architecture of a selection network.

the last control neuron fires, the right-most portion of the input pattern is selected. This type of selection network, however, can only select subimages in one direction—horizontal or vertical—not both. Thus according to our previous discussion, if this type of selection network is used in conjunction with the retinal–cortical transformation, then the retinal image can be processed for size standardization or rotation, but not both.

The question now becomes this: if the retina–cortical connections are responsible for size and rotation processing, then at what stage of visual processing does the selection take place? This will be the topic of the following discussion.

A POSSIBLE ROLE OF THE CEREBRAL CORTEX AND LATERAL GENICULATE BODIES IN VISUAL PREPROCESSING

At the level of processing of the primary visual cortex, the visual pattern is extremely large and consists of the million or so "picture elements" generated by the retina. A

FIG. 14. An input image with four different fixation points indicated. The cortical representations for each fixation point are indicated. Notice how the feature which is closest to the center of the fixation occupies a larger part of the cortical representation than surrounding features.

neural pattern of this size is much too large for permanent storage, and our face recognition experiments with digital images support the fact that it is also much larger than necessary for recognition. In fact, reasonable estimates for the number of neurons comprising stored visual patterns are under 1000 cells, and may be closer to three or four hundred. This suggests that a large-scale reduction in the size of the primary visual pattern occurs prior to storage, where the pattern is reduced from the primary sensory representation consisting of the million or so retinal ganglion cells to a secondary representation consisting of a few hundred cells. We believe that this gross size reduction occurs in the primary visual cortex in conjunction with size and rotation processing.

The primary visual cortex is conceived to be a visual sensory buffer whose architecture is shown in Fig. 16. Conceptually, this storage device is divided into independent

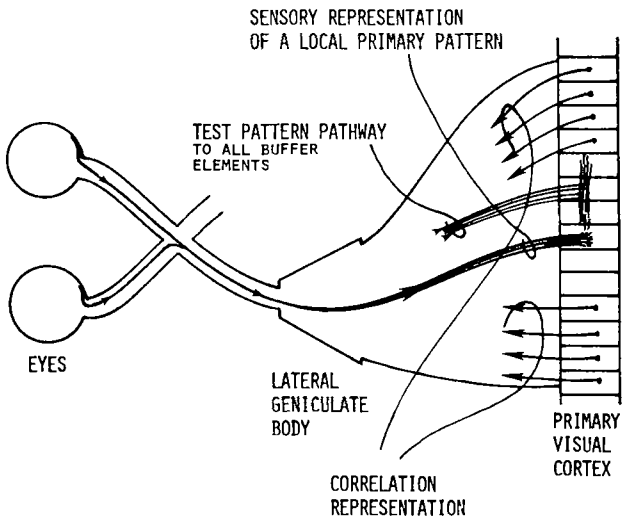


FIG. 16. The elementary visual preprocessing networks. The retina forms the primary representation of the visual world. The lateral geniculate bodies are implicated in pattern normalization and masking, and possibly in the formation of the test pattern and also in the selection of the secondary visual image for the focus of visual attention (see text). The primary visual cortex acts as a sensory buffer, but also correlates the local sensory patterns against stored images. This is the basis of object (region) extraction. The topology of the projections to the primary visual cortex also implicate it in size standardization and gross reduction of pattern dimensionality.

functional units called *basic computational elements* by Kabrisky (1966), but we will use the term *sensory buffer element* or *buffer element* for short here. Each buffer element is an associative storage location just as the memory locations are in a permanent experiential memory store.† We will suggest that the usual storage and recall control

† In fact, all indications are that the entire cerebral cortex is a storage organ, where different regions are involved with different types of storage. Some regions are permanent stores of experience, others are sensory buffers, others store premotor patterns, others store both verbal and non-verbal information and associate the two modalities. The structural differences between cortical regions are due to differences in storage parameters: temporary versus permanent traces, large versus small input patterns, short versus long storage time intervals, and so forth.

inputs to each buffer element are absent, but as illustrated in Fig. 16, there are two input patterns: the retinal pattern from a localized place on the retina, and a *test pattern* whose origin will be discussed later. The test pattern is distributed to all buffer elements just as the input patterns are in a permanent memory store.

One output from each buffer element is the correlation between the currently stored pattern in that buffer element and the current input pattern, either the local retinal input pattern to that buffer element or the test pattern. The set of all such correlation signals is a representation of the visual field which depends on the set of patterns stored in the buffer elements at the time, and the current input pattern (sensory pattern or test pattern). We shall call this representation of the visual field the *correlation representation.* The secondary visual pictorial pattern described earlier is obtained from the correlation representation by a selection network. As described earlier, the secondary pictorial patterns form the pictorial component of visual experiences and are therefore the basis of all high-level visual processing.

Based on these assumptions, the following questions naturally arise. What patterns are stored in the sensory buffer elements? What is the nature of the retinal input pattern to each sensory buffer element, and what is the nature of the test pattern? Finally, what is the origin of the test pattern? Although there are no definitive answers to these questions, we will suggest answers that are based on experimental and psychological evidence as well as information processing considerations.

The primary visual cortex is organized into columnar processing elements that receive inputs from the retina via the lateral geniculate bodies. These inputs, which arrive at the primary visual cortex in layers IVc, as well as I, IVa, and possibly VI, convey information about the nature of the retinal images in a relatively small region, including (but not exclusively) color, texture, and brightness information. These local primary patterns are immediately transformed by the cortical connections to a more complex representation of the visual field in layers II, III, V, and VI. Although there are numerous ways to represent visual information, our ability to choose which component we wish to attend suggests that a representation is used that enhances the selection processes. For example, we can look for an object with a specific color or texture, at a specific distance, or one which is moving. Marr (Marr, 1976; Marr & Poggio, 1977) calls this representation a *primal sketch,* but the question still remains, what is the nature of the transformation from retina to cortex and beyond?

One transformation, which appears to be particularly likely is the Hueckel (1971) transform, which is the product of radial and angular spatial Fourier transformations, but other transformations are also likely. In the visual system, the transformation would be made of local regions of the retina.† These local transformations have the property that the original image can be reconstructed from the transformed pattern and, in addition, they have other properties of special interest here. For example, one component represents the average light intensity in the corresponding region of the visual field. Another component represents the horizontal gradient of light, and yet another component represents the vertical gradient of light. Still another component represents the change in light intensity moving radially outward from the center of the region—the amount of dark center with light surround. These are *primitive features.*

† The connections of the retina, lateral geniculate bodies, and primary visual cortex suggest that local rather than global transformations are involved.

The classic studies by Hubel & Weisel (1968, 1977) confirm the existence of complex cells whose behavioral properties support this type of transformation.†

We now return to a discussion of the visual buffer, and assume that one of its functions is to perform such a local transformation of the corresponding pictorial image that it receives from the retina. These local transformations are then stored in the sensory buffer elements. The test pattern, which is distributed to all buffer elements, is the second pattern which can be stored by the sensory buffer elements.

What about the origin of the test pattern? We believe that there are several different origins whose choice at any given time is made by the visual selection and attention control networks. One origin is from a permanent visual memory store. When the test pattern comes from storage it may represent any prior feature that was once seen in the visual field. As a specific example, it might represent a primitive feature such as brightness, color, texture, horizontal gradient, and so forth. Each of these primitive features is immediately available in the transformed retinal patterns (see Hueckel, 1971). As stated earlier, the output of the sensory buffer is the set of correlation values between either the retinal patterns or a test pattern, and the patterns stored in the corresponding buffer elements. For this discussion we will assume that the patterns stored in the buffer elements are the transformed retinal patterns. If the test pattern representing the feature "brightness" is the input test pattern to the visual buffer, then the correlation output pattern is a pattern that represents the average light intensity in each local region of the visual field. The sensory buffer simply reduces the retinal pattern from the million or so cells in the optic nerve to a pattern whose size is equal to the number of columnar processing elements in the sensory buffer. This number is estimated to be a few thousand. The pattern itself represents the pattern of light intensities in the visual field, and therefore for this particular test patten the sensory buffer simply acts as a size-reduction network.

As a second example, if the test pattern represents the feature "x-gradient of light in the visual field", then the output pattern from the sensory buffer would correspond to the horizontal gradient of the visual field patterns, only reduced in size just as was the brightness pattern just described in the previous paragraph. In this case, the output would indicate exactly those regions of the visual field that are light on the left and dark on the right, or vice versa. This particular pattern would respond to vertical edges in the visual field. If the test pattern represents "dark center with light surround" (also called the Laplacian), the correlation pattern shows regions with rapidly varying grey levels. Figure 17 illustrates the Laplacian and x-gradient for a digitized image.

Perhaps more interesting is the possibility that arises when the origin of the test pattern is not from permanent memory at all but comes from within the visual field. In this case, each buffer element responds with an indication of the similarity between its stored pattern and a pattern derived from elsewhere in the visual field. If, for example, the subject is looking at an object and the test pattern comes from within that object, then the test pattern represents the color, texture, and brightness of the object. The resulting correlation signals are high wherever part of the object extends in the visual field and low where it does not. The pattern of high correlation signals is in the shape of the object. Simply stated, the sensory buffer when used in this way "extracts objects"

†It is clear that the primary cortical representation of the visual field contains depth (disparity) information as well as color, texture, and brightness information. It is also clear that depth information does not play a fundamental role in face recognition. Hence we consider only monocular visual information processing here.

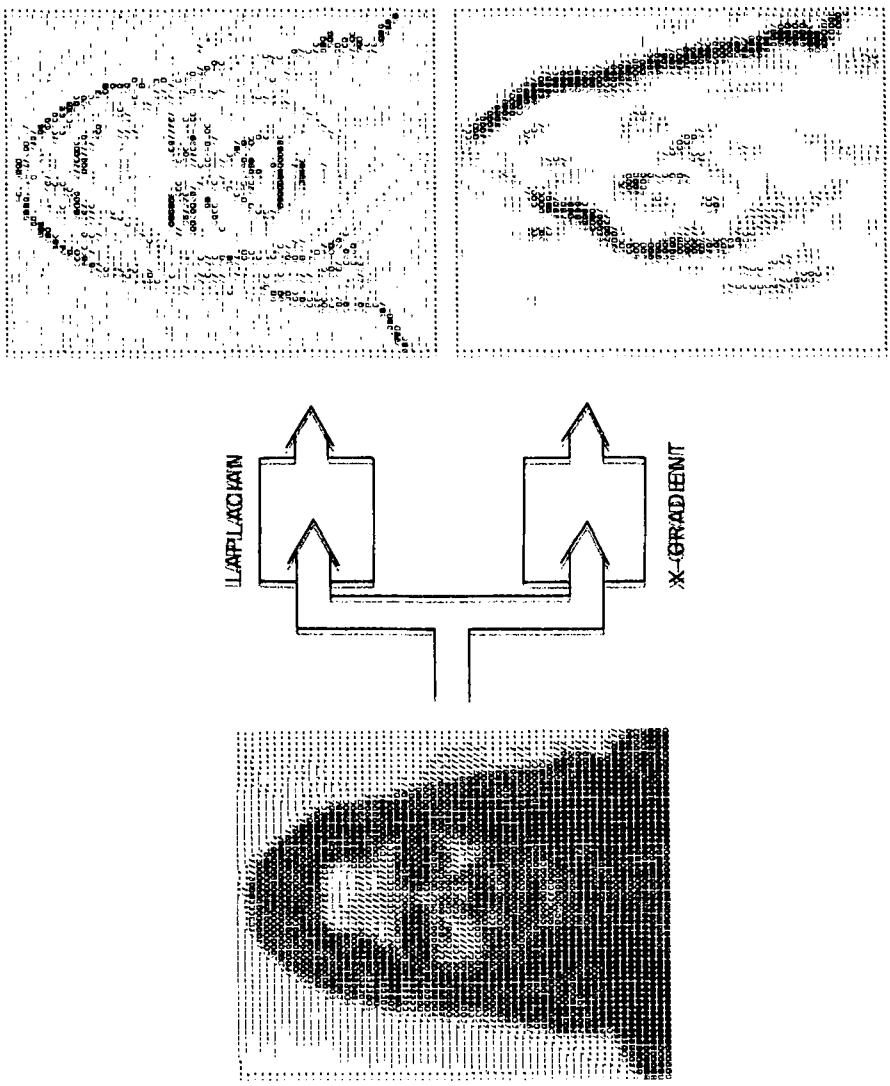


FIG. 127. The Laplacian and x-gradient of an input image.

from within the visual field. More precisely, it isolates regions of the visual field having a constant color or texture. The resulting correlation pattern can then be used as an "object mask" for background elimination. Networks for masking are not shown, but the process, which is illustrated in Fig. 18, can easily be accomplished in networks such as the lateral geniculate bodies, which relay information from one place to another.† [See Schiller (1977) for a related discussion, and also Turvey (1973) for a discussion of experimental studies of visual masking.]

In our earlier discussion we described how the cortex may produce a secondary pattern that is size and rotation invarient. Combining these two types of processing, it appears that the primary visual cortex in conjunction with the lateral geniculate bodies

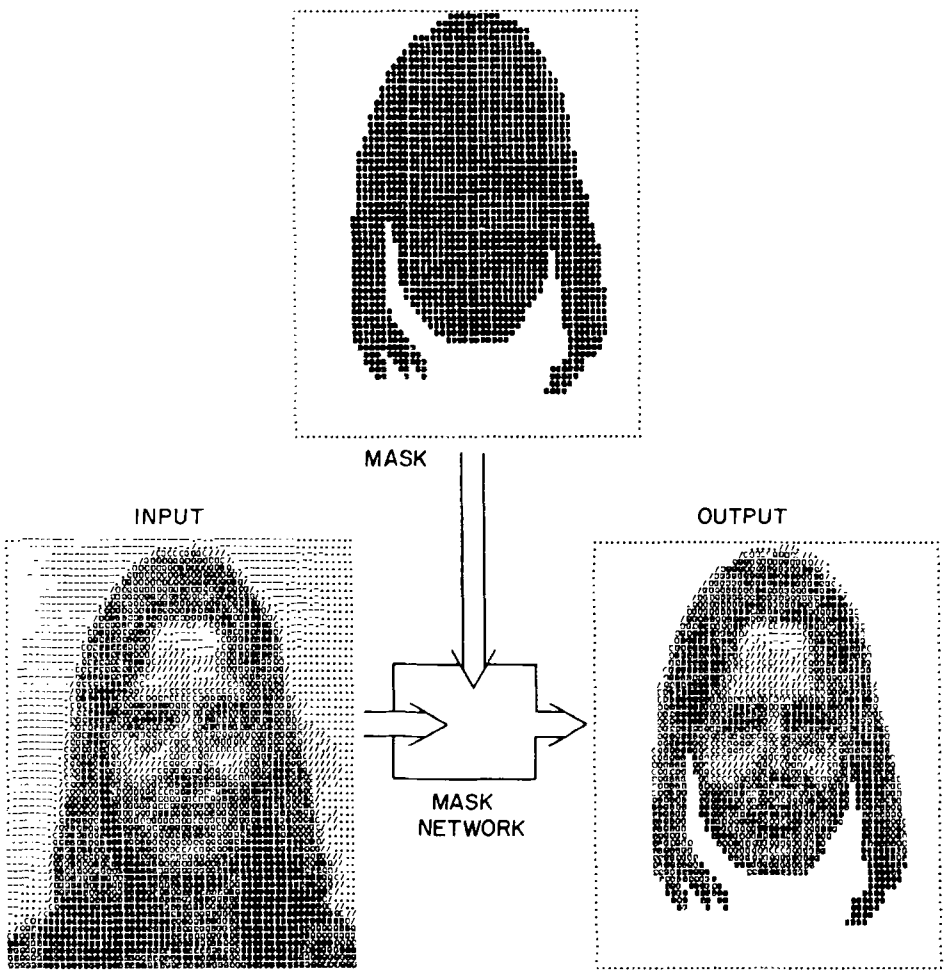


FIG. 18. An input image, a mask, and an output image formed by masking the input image.

† More important, images sent to a storage network must be normalized (Baron, 1970*a*). The lateral geniculate bodies can easily perform normalization, which is simply the uniform reduction or enhancement of firing rates of all cells in a local pattern so that the integrated output level is a specified constant.

perform several elementary visual processes: conversion of the retinal input pattern into an intermediate visual pattern that is size and rotation invarient, selecting an object defined by a particular attribute such as color, texture, or brightness (to name three), and reducing the dimensionality of the visual pattern from a million or so cells to a few thousand cells. The correlation pattern produced by the primary visual cortex then enters a selection network which selects the subimage that we have called the secondary pictorial representation of the visual field. The selection network may be the cortex, lateral geniculate bodies, or some other related neural structures.

The secondary pictorial pattern, combined with the control component that regulated its formation, including the test pattern used in its generation, and possibly a visual mask, is the representation of the visual field that forms our records of visual experience. This representation also forms the input to other networks for recognition, including face recognition. By including the test pattern in the storage representation, the storage representation includes not only the gross shape of the object (from the correlation pattern) but also its color and texture (from the test pattern component) and its size and location in the visual field (from its control component).

CONTROL OF THE SENSORY BUFFER

Unlike the permanent stores of visual experience which receive and store highly processed patterns that last for many seconds, a sensory buffer is continually bombarded by sensory inputs. As a storage device, a sensory buffer holds information for a relatively short time, perhaps a few tenths of a second, and hence the notion of storage and recall of information that is appropriate for a permanent memory store is not appropriate for a sensory buffer and must be re-evaluated. Although the specific nature of the storage control network for the visual sensory buffer is not known, what we suggest as one possibility will now be described. Input patterns that arrive either directly from a sensory pathway or through the alternate test pattern pathway are continually stored. According to our storage model, this means that the coupling parameters in the cells that effect storage are set according to the input pattern. The rate at which proportionality is achieved, of course, depends on the parameters of the sensory buffer, but a few hundred milliseconds is not unlikely in the visual system. Said another way, the memory traces are adaptive and continually change to match the current input pattern.† If the input pattern to the buffer does not change, then as soon as the sensitivity values adapt to that input pattern, each correlation output will be a maximum value. (The maximum correlation value is achieved when a pattern is correlated against itself.) Such a pattern conveys no information at all. If the input pattern suddenly changes, then the correlation output is the correlation between the old and the new patterns. However, as the memory traces once again adapt to the new pattern, the correlation output again approaches its maximum value.

The input pattern can change for several reasons. First, the control networks can switch the input from sensory pattern to test pattern or vice versa. Second, the sensory

† By analogy, consider a piece of photosensitive glass that turns dark when exposed to light. After a short time, those regions where the light is very bright become dark, and those where there is little light become clear. If an image is projected on the glass, the transparency of the glass will change according to the geometry of the projected image, and the glass will store a representation of the projected image. If the projector is turned off, the stored representation of the image will fade away. The glass continually adapts to the arriving light pattern.

pattern can change with a change in the fixation point of the eye. Finally, the test pattern can be altered by the control networks.

One possibility to consider is that under normal conditions the control networks continually shift between test and sensory patterns. As a result, the reduced output pattern may be thought of as a sequence of representations of the visual field. This is an example of a temporal sampling process. If a particular object is being attended, the sequence would be the result of presenting the entire sensory representation followed by presenting a test pattern derived from the center of the object. The pattern in the center of the object represents its color, texture, and brightness. As the attention shifts from object to object, the sequence of output patterns will be in the shape of the selected objects. The sensory buffer *segments* the visual field into regions having similar color and textural properties, and the segmented region changes as the attention shifts from object to object.

A second possibility is that the memory traces depend only on the sensory input patterns. That is, the sensory patterns cause the memory traces to be formed, but the sensory patterns do not directly influence the firing rates of the correlation cells. The test pattern, on the other hand, is correlated against the currently stored pattern, but does not affect the memory traces. In this case, there would be no need to alternate between sensory and test patterns since the correlation pattern would always represent the similarity between the stored patterns and the test pattern. The same is true, of course, if the test pattern effects storage and the sensory pattern does not.

DETECTING A FACE IN THE VISUAL FIELD

In order to place into proper perspective with other visual processes the operations described here, we will give a brief description of what we believe happens when a face is first seen.

The first step in any face recognition task is the detection of the face to be identified. Although the model presented here does not address this aspect of face recognition, a few words seem waranted. Before a face can be identified, it is absolutely essential that the visual processing networks locate the face and direct the eyes toward it. Networks that detect movement or change in the visual field (such as when a picture of a face is presented) initiate the orienting response that enable the face to be centered in the visual field. This is a reflex action which is essential to visual recognition. Once the face is located, the recognition processes described throughout this report can begin.†

The first step is the determination of size of the retinal image. This is most likely done in one of two ways. First, the correlation representation may be sequentially scanned. This is equivalent to changing the focal length of a zoom lens, only the zooming results here because of the topology of the retinal–cortical projections. This process would be terminated when the permanent visual memory stores respond with strong correlation signals, indicating the presence of a recognized object—in this case a face. Second, the attention networks may automatically select regions containing objects. An "object" is a region in the visual field that responds with high correlation signals when the test pattern is taken from within that region, and the selection and attention control

† It is perhaps no accident that the eye responds so readily to patterns having a dark center and light surround. The iris of the eye, eye-ball, eye socket, eyebrow, face, and hairline form a series of light and dark concentric circular regions that can easily be detected in the visual field.

networks simply isolate such regions in the correlation pattern and select them for the focus of visual attention.†

Once the preprocessing networks have extracted an image that is recognized by the permanent memory stores, subsequent processing can be directed by stored information. If the face is immediately recognized, which for a familiar person is most likely, the process may be terminated immediately. Otherwise, if stored information is available, the selection and attention networks can use the recalled control information to direct the scan of the visual field. The processing mode switches to verification, and the stored control patterns are used to direct the focus of attention from feature to feature in order to verify that the current face is a specific known face. At the same time, the stored test patterns can be used to reduce the sensory representation to the secondary representation. If the same color or texture test pattern is now used that was used when forming the stored representation of a known person's face, the resulting current secondary representation would only be the same if the textural and color properties of the unknown face are similar. Thus facial discrimination is possible not only based on shape but also on more subtle distinctions such as skin tones and complexion.

The control processes associated with recognition are complicated and include, among other things, the routing of the reduced sensory patterns to the proper storage networks, control of those storage networks both to store the current inputs and to make available for use the associated stored patterns, delivery of recalled patterns from storage to the selection and attention networks for use during verification, and so forth. In addition, the selection and attention control networks must regulate all the various perceptual processes such as switching between sensory and test pattern, selecting the proper region of the correlation pattern for transmission to the storage networks, and masking the visual patterns. Unless all of these processes are co-ordinated during all stages of processing, recognition is either impaired or does not occur at all.

The final sections of this paper discuss some of the consequences that damage to the networks described throughout this report would have on recognition (refer to Fig. 19 during the following discussion). The reader should consult one of the many excellent descriptions of prosopagnosia for a description of clinical impairments to facial recognition (Bornstein & Kidron, 1959; Cole & Perez-Cruet, 1964; Ellis, 1975; Gloning, Gloning, Jellinger, & Quatember, 1970; Levine, 1978; Meadows, 1974), or to any standard book on clinical neurology such as Luria (1966) or neuropsychology such as Walsh (1978) for descriptions of more general impairments of recognition.

Syndromes associated with the elementary visual processes

Figure 19 illustrates the major processing networks and many of the information pathways suggested by the model. These are summarized below.

GLOBAL CONTROL PATHWAYS AND THEIR FUNCTIONS

1. This pathway specifies the origin of the pictorial mask—either from permanent memory (C^3–A^2 or C^5–A^2), from the sensory buffer (B^3–A^2), or from the selection network (A^3–A^2).

†This process is computationally more difficult than zooming and requires much more neural circuitry than the zoom process mentioned earlier.

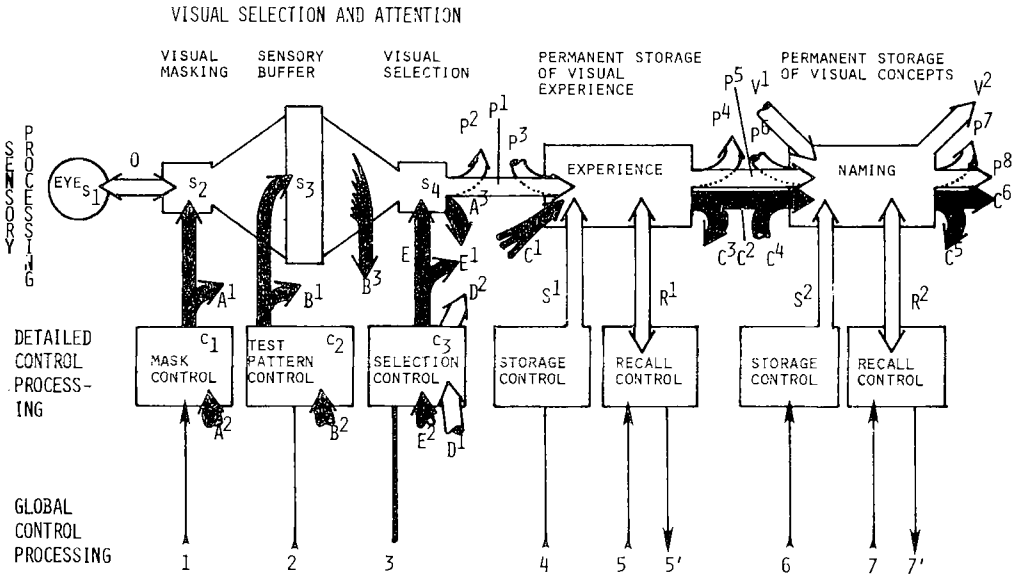


FIG. 19. The logical architecture of the major elementary visual processing networks (see text for an explanation).

2. This pathway determines the origin of the test pattern—from permanent memory (C^3–B^2 or C^5–B^2), or from the sensory buffer (B^3–B^2).

3. This pathway determines how the selection control pattern will be formed. The selection control networks co-ordinate the choice of secondary visual patterns with the positioning of the eyes, head, body and so forth (via the D^1–D^2 pathway). In short, this pattern specifies the focus of visual attention and is necessary for visual constancy.

4. This control pathway informs the storage control network when to store information.

5. This control pathway informs the recall control network when to recall information.

5′. This pathway delivers similarity information to the global control networks (not shown). Similarity information indicates when associated experiences are available for recall and can be used by the global control networks in deciding when to switch from undirected scanning to verification processing.

6. This pathway specifies when to store conceptual information—i.e. when to learn the name of an unknown person.

7. This pathway specifies when to recall conceptual information—i.e. enables access to the name of a known person provided that he or she is recognized.

7′. This pathway delivers similarity information from the naming store to the global control networks. This similarity information specifies when verbal information (e.g. the name of an object) is available for recall.

LOCAL CONTROL PATHWAYS

A. This pathway conveys the mask used for visual object extraction. A copy of the mask may be sent to a permanent storage network (thru A^1) as part of the representation of the visual experience. Pathway A^1 forms part of the pathways labelled C^1–C^6.

B. This pathway conveys the test pattern to the sensory buffer. A copy of the test pattern may be delivered to permanent storage (thru B^1) as part of the representation of the visual experience. Pathway B^1 forms part of the pathways labelled C^1–C^6.

E. This pathway delivers the selection control pattern to the selection network, and it therefore determines exactly what visual pattern occupies the focus of visual attention. A copy of this pattern may be sent to permanent storage (thru E^1) as part of the representation of visual experience. Pathway E^1 forms part of the pathways labelled C^1–C^6.

D. The pathways labelled D^1 and D^2 convey information about the control patterns that determine posture and position the eyes, head, body and so forth. The selection control networks must have knowledge of this pattern in order to co-ordinate the selection of visual field patterns as they change due to internally controlled movements.

S, R. The pathways labelled S^1, S^2, R^1, and R^2 are described in detail elsewhere (Baron, 1976).

SENSORY INFORMATION PATHWAYS

O. Optic nerve—the visual sensory input pathway.

P^1. Pathway to deliver pictorial information from the visual selection and attention networks to the permanent storage networks.

P^2. Pathway that delivers visual information to the motor system for visual–motor co-ordination.

P^3. Pathway to deliver recalled visual patterns that enable access to stored experiences. These patterns either come from experiential memory (P^4) when using one event to associate with another, or from the verbal processing networks (P^7) when accessing experiences verbally. ("Describe to me the movie *Close Encounters of the Third Kind*.")

P^6. Pathway to deliver pictorial inputs to the verbal processing networks. These inputs either come from experiential memory (P^5) when describing past events, or from the verbal processing networks (thru P^7).

V. Pathways V^1 and V^2 carry verbal inputs to and from the naming store.

PATHOLOGY

We will now describe how "damage" to the various networks and information pathways affects the behavior of the model, and although such a study does not verify its correctness, a close agreement between clinical pathology and corresponding pathology of the model does lend support to its correctness. The reader is directed elsewhere (Baron, 1976) for a similar discussion of the pathologies of the storage systems and their control networks.

In many complex systems, particularly the human brain, the networks can be loosely divided into three types: information pathways that connect one network with another, information processing networks whose behavior depends only on current and recent inputs, and storage systems. The information processing networks can be further divided into sensory processing networks and control networks, while the pathways can be divided into sensory pathways, local control pathways, and global control pathways. Local control pathways carry the detailed instructions that indicate exactly how the information processing networks are to process the current sensory patterns. Global control pathways, in contrast, regulate the gross behavior of the system—which

modality to attend, which storage network to use, where to deliver the processed sensory information, and so forth. The sensory pathways convey the brain's representation of the various sensory and motor events. The control networks process the global control patterns and determine precisely what action the information processing networks are to perform. For example, the global control networks may call for information from a permanent memory store whereas the recall control network, based on the recent history of correlation responses, generates a detailed control pattern to enable recall from a specific memory location.

Control may be exercised at several different levels of abstraction, and each network in the control hierarchy refines the specification until finally the precise action is determined. For economic reasons, this organization suggests that each level in the control hierarchy maintains autonomy over the networks that it regulates, but each higher level control network can monitor the outcome of each lower level process. For example, the global control networks can monitor the correlation signals from each storage network and therefore determine if the current inputs are recognized. The high-level control networks can therefore change processing strategy to suit the particular circumstances, and initiate a different behavior when necessary. As an example, recognition that a face is present in the visual field may indicate that a change from undirected search to verification is waranted.

For reliability, we suggest that the high level control functions and networks are located deep within the brain structures and are therefore much less vulnerable to damage from physical abuse whereas the sensory processing networks are closer to its surface and are therefore more prone to damage. Not coincidentally, the organization of the sensory processing networks appears to be highly parallel and modular so that damage to one part does not entirely disable its function but simply degrades its performance.† We note that replication of sensory function, modularization, and independence of local process appear to be three general principles of all sensory processing networks. Finally, we call attention to the architecture suggested by Figs 3 and 16 rather than Fig. 19. Figure 19 shows the connectivity but not the architecture, whereas Figs 3 and 16 illustrate how the input, output, and control pathways arrive perpendicular to the surface of the networks and are spatially distributed.

Damage to the visual selection and attention networks can occur in several different places, including damage to the processing networks themselves (at s^1–s^4), to the control networks (at c^1–c^3), to the various local control pathways, and to the global control pathways (at 1–3). We will describe possible syndromes caused by damage to the various networks and pathways.

SENSORY PROCESSING NETWORKS

Local damage to the sensory processing networks would cause visual field defects but would not defeat recognition. The extent of the impairment would be related to the extent of the damage, and damage to the central portions of the visual pathways would cause proportionately more degredation since they occupy a greater part of the system at higher stages of processing. Damage to the sensory buffer at places where the test pattern is formed would also cause a severe recognition impairment as it would degrade

† This may explain why evolution favored development of local transformations of the visual field rather than global transformations. If they were global, damage anywhere would render the system inoperative. Since they are local, damage only affects the damaged area, not the entire network.

region extraction and therefore object selection. Reading, which is a foveal process, may be spared. Damage to the secondary visual pathway at P^1 would impair recognition, and total severance of this pathway would cause total agnosia.

THE MASKING NETWORK

Masking is necessary when isolating figures which are embedded in complex images. For example, if one was to take a simple line drawing and scribble over it, most people would have no difficulty in recognizing the "hidden" figure. Gestalt recognition of a single part of the hidden figure would enable recall of an image of the figure which could then be used as a mask to eliminate the extraneous scribbles. Figures 7 and 8 illustrate this point. Damage to the pathways A or A^2 would degrade this ability, resulting in an inability of extracting simple figures from their background. Recognition of simple isolated figures may be spared.

Our computer studies of face recognition showed that distinctly different storage representations were formed for faces when photographed with and without flash. We suspect that within the brain, masking is used to compensate for the effects of shadows during recognition. In particular, when a face is lighted by sunlight, the eyes and the areas beneath the nose and mouth are darkened by shadows. The visual system can compensate for these shadows by enhancing the corresponding regions of the visual field. Such an enhancement operation can be performed in a masking network by using a mask which reduces the intensity of neural firing in all regions where shadows do not occur. It follows that damage to the masking network would result in difficulty in identifying faces when seen under different lighting conditions.

THE TEST PATTERN CONTROL NETWORK

Damage to the pathways that deliver the test pattern to the sensory buffer (B–B^3), or to the test pattern control network would have a severely degrading affect on the transformations that reduce the primary visual images into secondary images. If B were damaged, scene segmentation would be impaired and difficulty would be evident in forming representations of objects which appear in disjoint places in the visual field. For example, it would be difficult to appreciate the extent of a carpet which is partially hidden by many different pieces of furniture. Conceptual groupings of similar objects would also not be found—the pattern of similar flowers in a garden or similar books on a bookshelf.

If the test pattern pathway, B^1, from the visual selection and attention networks to the memory stores were damaged, color and texture recognition would be impaired even though object recognition would be intact. This should not impair the ability to discriminate between colors or textures when presented side by side in the visual field, or to use logical (verbal) inference properly (e.g. a banana is yellow), but rather it would impair the ability to identify (name) colors or their use (e.g. when shown a yellow card to specify that it is the color of a lemon). Difficulty would also result in identifying textures (e.g. the fuzzy texture of a peach). The question of whether color and texture recognition can be impaired independently depends on how the test patterns are generated and what specific representations are used of visual information. This is a problem for future research.

The impairment resulting from damage to the pathway, B^2, from memory would be substantially different. Test patterns from memory are used in forming the secondary

pictorial representation of the visual world, and loss of this ability would mean that recognition of objects based on subtle distinctions would be impaired. For facial recognition, this suggests that differences in appearance due to skin tones or complexion could not be directed from memory. It would be difficult, for example, to distinguish between two faces with similar shape. This would be like recognition from line drawings only, without being able to use stored textural or color properties to drive the encoding process.

Another resulting impairment would be an inability in locating objects with a specified color. For example, if told to find a red book on the book shelf, a person would have to look from book to book sequentially to determine its color. A person without this damage can easily detect an object with a specified color. The attention mechanisms can be "tuned" to the presence of the named color, and low level reflex processes direct the attention toward objects having the specified color. With damage to the test pattern pathway from memory, this ability would be impaired since the test patterns which encode color information could not be obtained from the storage system.

THE SELECTION CONTROL SYSTEM

Damage to the selection control system and related pathways would be most severe and result in the inability to use spatial relationships. Damage to the control pathway, E, would in all likelihood render the subject incapable of locating objects, selecting them for the focus of attention, or recognizing them. This would be so even though the visual field would remain intact. Selection would be reduced to some fixed region on the retina, and any changes in eye position, in subject to object distance, and so forth would drastically change the secondary visual patterns. Recognition would occur by accident if at all, and only if the secondary visual image just happened to be a known encoding of an object.

Damage to the pathway E^1 would result in an inability to name and use spatial relationships. Two side-by-side objects might be recognized, but the fact that one of them is "to the left of" the other could not be determined. Damage to the control pathway, E^2, from memory would have a similar affect. Here, recognition of complex images would be severely impaired since the subject could not direct the attention from object to object in the visual field. A familiar room would no longer appear familiar even though the individual items of furniture are recognized (in the Gestalt sense). The verification mode of visual processing would likewise be impaired, and recognition of complex objects would be based on the recognition of independent features, each of which can be recognized in the Gestalt sense. The ability to distinguish between similar faces would be impaired since sequential analysis of facial features is required.

Damage to the selection control networks would result in an inability to co-ordinate head, eye, and body movements with the selection process, and the severity would depend on how badly the networks were damaged. Visual constancy would no longer be maintained, and the result would be a chaotic and unmanageable visual world representation.

General comments

It is very doubtful that any one of the syndromes mentioned above would demonstrate itself in its pure form. Although the model suggests independent networks for the

various processes, the functions can easily by combined within a single network. For example, both selection and size standardization can take place within a network such as the cortex, and storage as well as the storage control functions can likewise take place within a single neural system. Moreover, the various pathways are not likely to be small isolated bundles of nerve fibers, but rather, they are likely to be intermingled with other pathways, much like the nerve fibers in the optic chaism. Figures 3 and 16 illustrate how the nerve fibers are likely to intermesh, so that damage to a single pathway is not likely to occur without damaging other pathways as well. The result, of course, is a combination of syndromes without implicating a specific functional network.

Syndromes of facial recognition (e.g. prosopagnosia) can arise in many different ways as explained in the previous sections. In addition, any damage to the storage systems or their control networks can also result in difficulty with facial recognition. It should be clear that within the experiential memory stores are the engrams that encode people's faces, so that damage to these networks would result in an inability to recall events associated with individual people. If all engrams pertaining to an individual were lost, then he or she would appear unfamiliar, as if never seen before. The same would be true, of course, if because of damage, the preprocessing networks form visual representations of faces that are no longer similar to prior stored representations. In this case, recognition would not be possible, and we suspect that many demonstrations of prosopagnosia are of this nature.

Our ability to name faces as well as recall events about people in response to questions about them suggests that in addition to experiential representations of faces are conceptual representations of individual people's faces, classes of faces, and facial features. Some of these conceptual representations are found in the visual naming stores. All else remaining equal, an inability to name faces is most likely to be due to specific control parameters and strategies learned and used by the naming store control networks, and damage to these networks can easily result in anomia even though the individuals are easily recognized and relevant experiential information available for recall.

Although it is possible that specific (specialized) networks are used for facial recognition, there is little evidence to support such an hypothesis (Ellis, 1975). Rather, it seems more likely that specific control procedures are learned for facial recognition, and these procedures are invoked when a face is recognized in the visual field. If the sensory representations change due to brain damage, then in addition to the other difficulties described above, access to these control procedures would be impaired and facial recognition would be made more difficult still. It is most likely that prosopagnosia is the result of a combination of both sensory and control difficulties rather than damage to networks whose purpose is facial recognition *per se*.

Concluding remarks

This paper has suggested several different networks and mechanisms that underlie facial recognition and, more generally, visual perception. We have described several of the logical operations that we believe underlie facial recognition, and we have proposed several neural networks for performing the operations. We have suggested a correspondence between some of the underlying neural networks and specific networks of the brain, but the correspondence is far from certain or complete. Although the

proposed networks account for many clinical syndromes and experimental observations, numerous questions remain unanswered. Among the more notable are the relationships between the mechanisms described here and our ability to manipulate mental images, and our ability to read, which represents an extremely fast sequential process of the foveal system. We have also omitted any discussion of the differences between the functions of the two cerebral hemispheres or how they interact. The model presented here is a model for part of the networks of a single hemisphere operating in isolation. Finally, we have omitted any in-depth discussion of the global control processes. Our model is simply one step toward understanding the mechanisms of visual perception, and we hope that it suggests avenues for further theoretical analysis as well as detailed anatomical and experimental investigations.

References

BARON, R. J. (1970*a*). A model for cortical memory. *Journal of Mathematical Psychology*, **7**, 37–59.

BARON, R. J. (1970*b*). A model for the elementary visual networks of the human brain. *International Journal of Man–Machine Studies*, **2**, 267–290.

BARON, R. J. (1974*a*). A theory for the neural basis of language. Part 1: A neural network model. *International Journal of Man–Machine Studies*, **6**, 13–48.

BARON, R. J. (1974*b*). A theory for the neural basis of language. Part 2: Simulation studies of the model. *International Journal of Man–Machine Studies*, **6**, 1.

BARON, R. J. (1976). Brain architecture and mechanisms that underlie language: An information-processing analysis. *Annals of the New York Academy of Sciences*, **280**, 240–256.

BARON, R. J. (1979). A bibliography on face recognition. *The SISTM Quarterly Incorporating the Brain Theory Newsletter*, **II**(3), 27–36.

BEAUDET, A. & DESCARRIES, L. (1978). The monoamine innervation of rat cerebral cortex: synaptic and nonsynaptic axon terminations. *Neuroscience*, **3**, 851–860.

BORNSTEIN, B. & KIDRON, D. P. (1959). Prosopagnosia. *Journal of Neurology, Neurosurgery, and Psychiatry*, **22**, 124–131.

CAREY, S. (1978). A case study: face recognition. In *Explorations in the Biology of Language*, Walker, E., Ed., pp. 175–201. Montgomery, Vermont: Bradford Books.

CAREY, S. & DIAMOND, R. (1977). From piecemeal to configurational representation of faces. *Science*, **195**, 312–314.

COLE, M. & PEREZ-CRUET, J. (1964). Prosopagnosia. *Neuropsychologia*, **2**, 237–246.

CORCORAN, D. W. J. & JACKSON, A. (1977). Basic processes and strategies in visual search. In *Attention and Performance VI*, Dornic, S., Ed., pp. 387–411. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

DICK, A. O. (1976). Spatial abilities. In *Studies in Neurolinguistics*, Vol. 2, Whitaker, H. & Whitaker, H. A., eds. pp. 225–268. New York: Academic Press.

ELLIS, H. D. (1975). Recognizing faces. *British Journal of Psychology*, **66**, 409–426.

GLONING, I., GLONING, K., JELLINGER, K. & QUATEMBER, R. (1970). A case of "Prosopagnosia" with necropsy findings. *Neuropsychologia*, **8**, 199–204.

GOULD, J. D. (1973). Eye movements during visual search and memory search. *Journal of Experimental Psychology*, **98**, 184–195.

HAILINE, L. (1978). Developmental changes in visual scanning of face and non face patterns by infants. *Journal of Experimental Child Psychology*, **25**, 90–115.

HARMON, L. D. (1971). Some aspects of recognition of human faces. In *Pattern Recognition in Biological and Technical Systems*, Grusser, O.-J., Ed., pp. 196–219. New York: Springer-Verlag.

HARMON, L. D. (1973). The recognition of faces. *Scientific American*, **229**, 70–82.

HOWE, J. A. M. (1970). Eye movements and visual search strategy. *Memorandum MIP-R-69*, Department of Machine Intelligence and Perception, University of Edinburgh.

HUBEL, D. H. & WIESEL, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, **195**, 215–243.

HUBEL, D. H. & WIESEL, T. N. (1977). Functional architecture of macque monkey visual cortex. *Proceedings of the Royal Society of London B*, **198**, 1–59.

HUECKEL, M. H. (1971). An operator which locates edges in digitized pictures. *Journal of the Association for Computing Machinery*, **18**, 113–125.

JUST, M. A. & CARPENTER, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, **8**, 441–480.

KABRISKY, M. (1966). *A Proposed Model for Visual Information Processing in the Brain.* Urbana, Illinois: University of Illinois Press.

KRIEG, W. J. S. (1966). *Functional Neuroanatomy.* Evanston, Illinois: Brain Books.

LEE, D. N. (1978). The functions of vision. In *Modes of Perceiving and Processing Information.* Pick, H. L., Jr & Saltzman, E., Eds, pp. 159–170. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

LEVINE, D. N. (1978). Prosopagnosia and visual object agnosia: a behavioral study. *Brain and Language*, **5**, 341–365.

LURIA, A. R. (1966). *Higher Cortical functions in Man.* New York: Basic Books.

MACK, A. (1978). Three modes of visual perception. In *Modes of Perceiving and Processing Information.* Pick, H. L., Jr & Saltzman, E., Eds, pp. 171–186. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

MACKWORTH, N. H. & MORANDI, A. J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics*, **2**, 547–552.

MARR, D. C. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London B*, **275**, 483–524.

MARR, D. C. & POGGIO, T. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Research Program Bulletin*, **15**, 470–488.

MAURER, D. & SALAPATEK, P. (1976). Developmental changes in the scanning of faces by young infants. *Child Development*, **47**, 523–527.

MEADOWS, J. C. (1974). The anatomical basis of prosopagnosia. *Journal of Neurology, Neurosurgery, and Psychiatry*, **37**, 489–501.

NEISSER, U. (1964). Visual search. *Scientific American*, **215**, 94–102.

NEWELL, A. (1973). Production systems: models of control structures. In *Visual Information Processing*, Chase, W. G., Ed, pp. 463–526. New York: Academic Press.

NOTON, D. (1969). A proposal for serial, archetype-directed pattern recognition. *Record of the* 1969 *IEEE Systems Science and Cybernetics Conference*, Philadelphia, 22–24 October, pp. 186–191.

NOTON, D. & STARK, L. (1971*a*). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, **11**, 929–942.

NOTON, D. & STARK, L. (1971*b*). Scanpaths in eye movements during pattern perception. *Science*, **171**, 308–311.

PRESTON, K., JR (1965). Computing at the speed of light. *Electronics*, **38**, 72–83.

PRIBRAM, K. H., NUWER, M. & BARON, R. J. (1974). The holographic hypothesis of memory structure in brain function and perception. In *Contemporary Developments in Mathematical Psychology, Vol. II. Measurement, Psychophysics, and Neural Information Processing*, Krantz, D. H., Atkinson, R. C., Luce, R. D. & Suppes, P., Eds, pp. 416–457. San Fransisco: W. H. Freeman.

PRINZ, W. (1977). Memory control of visual search. In *Attention & Performance VI*, Dornic, S., Ed., pp. 441–462. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

SCHILLER, P. H. (1977). The cortical and retinal inputs to the superior colliculus. *Neurosciences Research Program Bulletin*, **15**, 434–436.

SCHNEIDER, W. & SHIFFRIN, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, **84**, 1–66.

SCHWARTZ, E. L. (1977*a*). The development of specific visual connections in the monkey and the goldfish: outline of a geometric theory of retinotopic structure. *Journal of Theoretical Biology*, **69**, 655–683.

SCHWARTZ, E. L. (1977*b*). Afferent geometry in the primate visual cortex and the generation of neuronal trigger features. *Biological Cybernetics*, **28**, 1–14.

SENDERS, J. W., FISHER, D. F. & MONTY, R. A. (Eds) (1978). *Eye Movements and the Higher Psychological Processes.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

SHIFFRIN, R. M. & SCHNEIDER, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attention, and a general theory. *Psychological Review*, **84**, 127–190.

SPERLING, G. & MELCHNER, M. J. (1978). The attention operating characteristic: Examples from visual search. *Science*, **202**, 315–318.

TURVEY, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, **80**, 1–52.

WALKER-SMITH, G. J., GALE, A. G. & FINDLAY, J. M. (1977). Eye movement strategies involved in face perception. *Perception*, **6**, 313–326.

WALSH, K. (1978). *Neuropsychology.* New York: Churchill–Livingstone.

WEIMAN, C. & CHAIKEN, G. (1977). Logarithmic spiral grids for image processing and display. *Technical Report TR*77–3 (15 June). Old Dominion University.