

Sequence Seeking and Counter Streams: A **10** Model for Information Flow in the Visual Cortex

In this chapter computational considerations discussed throughout this book, combined with psychophysical and biological data, are used to propose a model for the general flow of information in the visual cortex. The model uses a process called "sequence-seeking" which is a search for a sequence of transformations and mappings that link an input image with a stored object representation. This process has two main characteristics: it is bi-directional, bottom-up as well as top-down, and it explores in parallel a large number of alternative sequences.

From a biological standpoint, this computation is performed by a structure called "counter-streams." This structure is composed of two complementary pathways, an ascending one from low to high visual areas, and a descending one going in the opposite direction, from high to low visual areas. Roughly, the suggestion is that the ascending pathway performs bottom-up processing, starting from the image and proceeding to high-level visual areas. The descending pathway is performing top-down visual processing, starting with stored models in higher level visual areas and proceeding to lower regions. Within each pathway, multiple processing sequences are explored. The integration of bottom-up with top-down processing is achieved by the interactions between the two complementary processing streams. A biological embodiment of this model in cortical circuitry is proposed. The model serves to account for known aspects of cortical interconnections and to derive new predictions.

The focus of the model is on the task of visual recognition and the structure of the visual cortex. The proposed computation has, however, useful generic aspects, and the possible applicability of the scheme to other domains will also be briefly considered. The first part of this chapter will focus on the computations being performed, and the second on details of the biological model. The second part assumes some familiarity with the principle features of cortical anatomy and physiology. Good summaries of the biological background can be found in (Crick & Asanuma 1986, Hubel 1988, Zeki 1993).

10.1 The Sequence-seeking Scheme

As we have seen throughout this book, object recognition is complicated by the large variations that exist between different images of the same object. To obtain effective recognition, it is therefore insufficient to perform a direct matching between the input and the stored patterns. Considerable processing is required to compensate for the effects of viewing direction, illumination, occlusion, and deformations in the object itself. The discussion in previous chapters led to several conclusions regarding the general nature of the processes involved in object recognition, including the use of pictorial representations, the combination of universal bottom-up with top-down, object and class-specific processes, the establishment of a correspondence between the image and the stored model, and the role of classification on the way to individual identification. These conclusions will be used below to motivate key aspects of the proposed model.

10.1.1 Bi-directional Processing

The approach to recognition developed in the previous chapters includes as a major component the manipulation and matching of pictorial representations. This matching is complicated by the fact that the image and model representations can be initially quite dissimilar. The recognition process therefore includes processes that compensate for the initial discrepancy between the two representations. These compensation processes make use of stored information based on accumulated past experience, at both the object and class levels.

In most of the alternative approaches to recognition, including the invariant properties scheme as well as the structural description method, compensation is obtained without consulting the stored internal models: recognition proceeds by processing the incoming image to obtain a new representation that is invariant to the effects of viewing conditions, such as illumination and viewing direction. These representations are matched at a final stage with

object models stored in memory. In contrast, in the schemes discussed earlier (including 3-D alignment and the view-combination methods) the stored models play a much more active role. For example, the compensation for the 3-D viewing direction does not require the extraction from the image of a 3-D invariant description. Instead, the compensation is based primarily on the use of 3-D information stored with the object model, either at the level of the object's general class, or at the level of the specific object model. Similarly, effects of object deformations, such as facial expressions, are compensated for by applying class-specific processes, as opposed to some universal bottom-up processes. In the case of illumination effects, which are often assumed to be handled by purely bottom-up processes, evidence was also cited supporting the view that they are compensated in part by class-specific processes adapted, for example, to the class of upright faces.

The use of stored information in the recognition process can take various forms. For example, to compensate for the effect of viewing direction, the method of 3-D alignment uses the explicit manipulation of an internal 3-D model, while the image combination method uses for the same purpose combinations of stored views. The crucial point for the current discussion is not the particular method used, but the basic strategy of using stored object- and class-specific information. We have seen that the compensation for viewing conditions on the basis of the image alone is difficult, and may require the extraction of precise 3-D shape as well as the illumination conditions. Instead of relying exclusively on bottom-up processing of the input image, the recognition process can be aided by using the results of accumulated past experience with the same or similar objects. For example, the human visual system is repeatedly exposed to face images, seen from different directions and under a variety of illumination conditions. It would be clearly advantageous to use some general learned properties of faces in the recognition of a novel face image. These properties include the 3-D shape of a typical face, the effects of illumination, shadows, and specularities, as well as common facial expressions.

Bottom-up and Top-down Processing We have already used above and in previous chapters the notions of “bottom-up” and “top-down” processing, which are frequently used terms in the psychology of perception as well as in computer vision. The distinction is that bottom-up processes are involved in the analysis of the incoming image, and top-down processes originate with stored models and information associated with them. The term “top-down” is sometimes used in a more restricted sense, to refer to the influence on perception of expectations and high-level contextual information; for example, the kind of knowledge one can bring to bear when one is entering a familiar office. Here, the term is used in a broader sense, to refer to processing applied to the stored models rather than the incoming image. Such processes can use knowledge about individual objects and object classes in the recognition process, including information concerning 3-D shape, the effect of illuminations, and deformations that an object may undergo.

In computer vision, the integration of bottom-up with top-down processing has been a major concern. During the 70's, the emphasis was placed heavily on top-down processes. Fundamental difficulties with building computer vision systems led to the view that the processing must be guided primarily by knowledge associated with stored models of objects and scenes, and systems in the 70's were constructed using this approach (Freuder 1974, Tenenbaum & Barrow 1976). Following in part the work of Marr (1982), the emphasis shifted towards bottom-up processing, but it became also evident that a key issue is the integration of bottom-up with top-down processing (Grimson 1990a, Marr 1982, Tsotsos 1990), and both processes are used in current recognition systems (e.g., Grimson 1990a, Lowe 1985, Ullman 1989, Yuille & Hallinan 1992).

In many neural network models the emphasis has been placed on bottom-up processing. Networks of the back-propagation type, for example (Rumelhart, Hinton & Williams 1986, Matan *et al.* 1992) are essentially feed-forward networks, where the computation pro-

ceeds in one direction from the input layer through successive intermediate stages to the output layer. The flow of information in the opposite direction is used for a different purpose—modifying the network when the errors in its output become too large. The general view expressed by Hubel and Wiesel in describing their findings regarding the visual cortex (Hubel & Wiesel 1962, 1968) also emphasizes the bottom-up direction. This informal model regards visual information processing as the successive extraction of increasingly elaborate image features. Different cell types in the visual cortex, in particular simple, complex, and hyper-complex V1 units, were originally thought to form the first stages of this hierarchy. Some neural network models, such as the Neocognitron (Fukushima 1988), attempted to simulate this mode of processing and apply it to pattern recognition problems.

In contrast with this approach, the computational studies examined in previous chapters argue in favor of a more balanced combination of bottom-up and top-down processing. As we have seen, the processes of object recognition, including the difficult tasks of compensating for the effects of viewing direction, illumination, occlusion and object deformation, can benefit from the use of information associated with objects and object classes, acquired through past experience. A major requirement for a model of information processing in the visual cortex is therefore the capacity to combine efficiently bottom-up processing starting at the image and proceeding to high-level cortical areas, with top-down processing, starting at stored object representations, and proceeding from high to low visual areas. As will be discussed in more detail later, the structure of the cortex clearly supports this possibility. A major characteristic of cortical interconnections is the reciprocity of the connections between visual areas. If a visual area in the cortex sends ascending connections to another visual area higher up in the hierarchy of visual processing, then, as a general rule, the second area sends reciprocal connections to the first. (This connectivity, including the so-called lateral interconnections, will be examined in more detail in the second part of this

chapter.) A key ingredient of the proposed model is the suggestion that, roughly speaking, the ascending pathways in the visual cortex subserve mainly bottom-up processing, and the descending pathways mainly top-down processing. Certain interactions between the streams provide a mechanism for integrating the two types of processing.

It is worth noting that in the matching of pictorial descriptions discussed in previous chapters, the use of bi-directional processing, top-down as well as bottom-up, is considerably more natural than in alternative approaches to recognition, because similar processes can be applied to the incoming image and to stored object models. This can be contrasted, for example, with schemes based on the use of invariant properties, discussed in chapter 2. Bottom-up processes are used in these schemes to extract from the image a vector of property values, but it is not apparent how to proceed in the opposite direction, that is, how processing could be applied to the property vector in a top-down fashion to help the recognition process.

10.1.2 Exploring Multiple Alternatives

A second general strategy proposed by the current model is based on a combination of computational considerations and biological constraints. Computational experience with object recognition schemes, including 3-D alignment and the view-combination method, have shown that the recovery of the appropriate compensation transformations often require considerable search. For example, some of the best-performing recognition systems using 3-D alignment (Fischler & Bolles 1986, Grimson 1990a, Huttenlocher & Ullman 1990, Lowe 1985, Thompson & Mundy 1987), search for the best match between an input shape and a candidate internal model by exploring and comparing multiple (e.g., in the hundreds) possible 3-D poses of the internal model. As a simple illustration of the problems involved, suppose that to compensate for possible size variations, we wish to normalize the input image in size, that is, scale it either up or down, to bring it into agreement with a

predetermined canonical size. One problem that arises is how do define and measure the size of the input image. Typical measures used in practice include the apparent area of the object in the image, or the size of the smallest shape (such as a square or ellipse) containing the viewed object. Such measures can give an estimate of the object's size, but a precise value is difficult to obtain. For example, when applied to a face image, the measured size will change with changes in viewing direction, will be affected by shadows and occlusion, will change with the hairdo, and so on. This is a simple example of an ubiquitous problem: the alignment process cannot be performed instantly and uniquely, since a search for the appropriate parameters is usually unavoidable. This may seem a somewhat practical consideration, but it has significant implications due to the inherent slowness of neuronal systems.

In current computer systems the multiple comparisons are performed sequentially, at a rapid succession. In much slower neuronal networks, timing considerations (Maunsell & Gibson 1992, Rolls, Tovee & Lee 1991, Thorpe *et al.* 1991) place rather stringent restrictions on the use in recognition of long chains of calculations, or the use of iterative relaxation processes (Geman & Geman 1984, Poggio, Torre & Koch 1985, Mumford 1992). A visual cortical area may introduce an average delay of about 10-15 milliseconds, and there are several (about six) stations spanning the hierarchy from V1 to anterior IT. This suggests that visual processing should usually require a limited number of sweeps through the system.

Neuronal systems are slow, but inherently parallel, and therefore a useful strategy is to explore multiple alternatives simultaneously, rather than explore and refine them in sequence. In some widely used neural network models, such as back-propagation (Rumelhart, Hinton & Williams 1986) or the Hopfield model (Hopfield 1982), the simultaneous computation of competing alternatives is usually not used explicitly. It is interesting to note, however, that in several neural network models in areas such as arm control (Jordan & Jacobs 1993), or handwriting recognition (Matan *et al.* 1992), the explicit exploration of multiple alterna-

tives proved useful in dealing with complex problems that were not handled effectively by more standard models. For example, in a network developed at AT&T for reading hand-written numerals (Matan *et al.* 1992), the letter identification stage is preceded by a process that segments the string into individual characters. This turned out to be a difficult task, and the segmentation was often unreliable. The approach adopted was to perform explicitly a number of different candidate segmentations, and later select the best solution among the different alternatives. The network also uses multiple-scale analysis: it processes the same input through separate mechanisms tuned to different scales of the letters, and again selects at the end the best alternative. This design proved to perform better than a single network trained to become scale-invariant. Another example is a network developed for controlling the 3-D movements of a simplified arm (Jordan & Jacobs 1993). The training of standard neural networks indexNeural network models failed to converge to a satisfactory solution to this control problem. An approach that proved successful was to train a number of different sub-networks, each one capable of providing a good solution under restricted conditions. The combined problem is then treated by letting the different networks work on the problem individually, followed by a gating and selection stage.

10.1.3 The Counter-streams Structure

The main conclusions reached so far are illustrated pictorially in figures 10.1 and 10.2 in the domain of face recognition. Figure 10.1a illustrates the problem: the discrepancy between the stored model and the novel image is initially large, and the two cannot be matched directly. Figure 10.1b illustrates the combination of bottom-up and top-down processing. The processes compensating for the image-to-model differences are applied in part to the stored model and in part to the incoming image. Top-down processing is involved in dealing with the effects of viewing direction, facial expression, and illumination. Bottom-up processing is used in the figure to handle variations due to position and scale. The

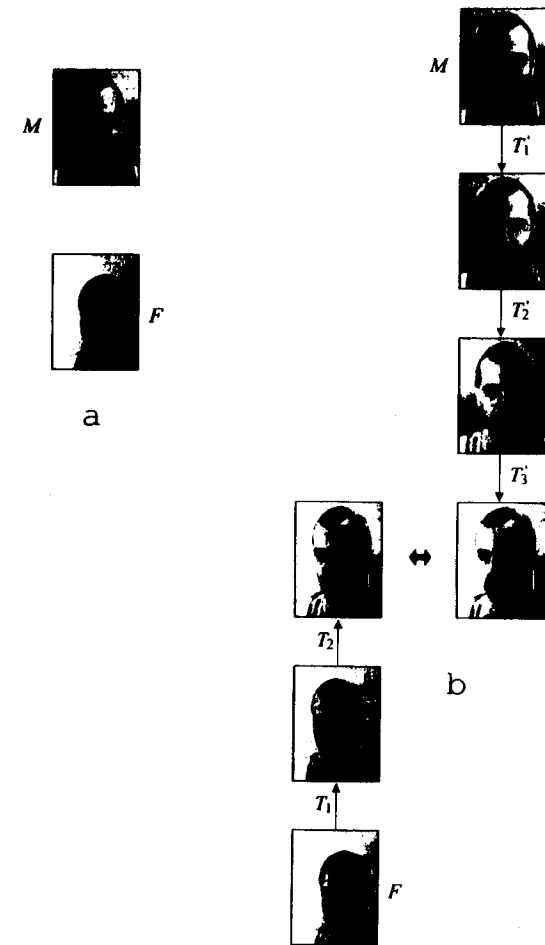


Figure 10.1

a. The stored model and the input image are initially different, and cannot be matched directly. *M* represents the stored model, *F* is the novel image. *b.* During the recognition process, bottom-up processes are applied to the incoming image, and top-down processes are applied to the stored model.

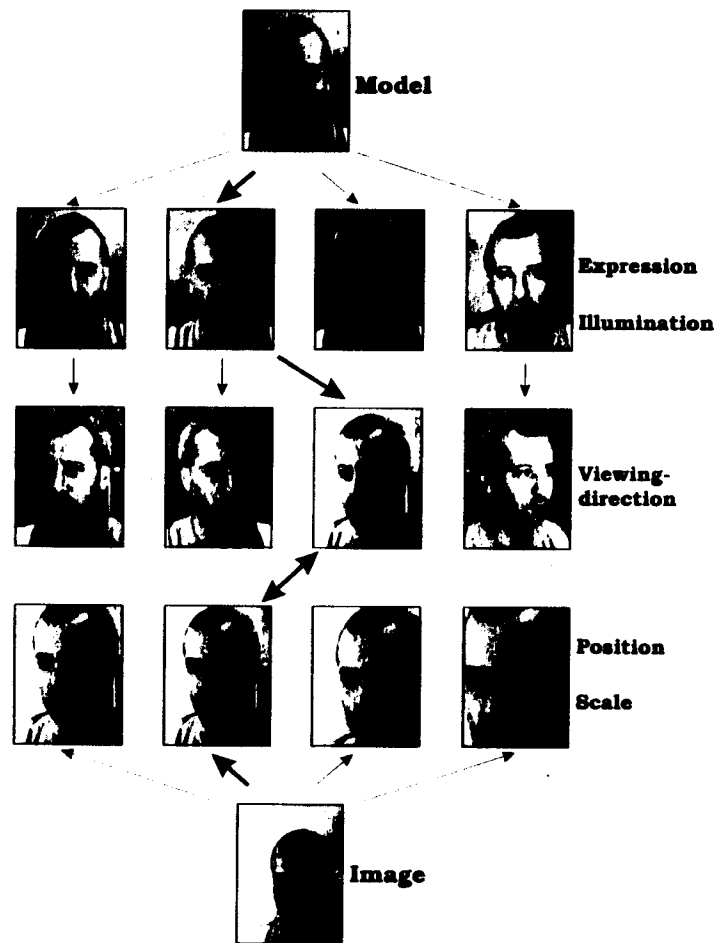


Figure 10.2
Illustration of two main properties of the sequence-seeking process: it uses bi-directional processing, and the simultaneous exploration of multiple alternatives. The top image represents a stored face model, and the bottom image is the novel view to be matched with the model. In this example, scale and position are processed by bottom-up processing; facial expression, illumination, and viewing position, by top-down processing.

model is depicted as a picture to represent the use of pictorial descriptions in the recognition process. However, the stored model is actually not just a grey level image, but a more abstract pictorial representation that also contains more than a single view.

In this figure, a single stream of processing is applied from both directions. How are the appropriate transformations selected at each stage by the system? As discussed above, this is obtained by evaluating in parallel multiple competing alternatives. Figure 10.2 illustrates the exploration of multiple alternatives in the two processing streams. The image at the top represents again a stored internal model of a face, and the image at the bottom is the novel input to be matched with the model.

The novel view differs from the stored model in viewing direction, illumination, and facial expression, demonstrating the limitation of direct image matching. The figure illustrates schematically the two properties of the compensation process discussed so far: the bi-directional processing and the exploration of multiple alternatives. Processes applied to the image are involved, for example, in compensating for differences in position and scale. This is represented in the figure by multiple internal images, at different positions and scales, generated in a bottom-up manner on the ascending stream of processing. Processes applied to the stored model are used in handling the effects of facial expression, illumination, and viewing position. This is represented by multiple internal images, at different illumination and viewing directions, generated from the stored model on the descending stream. The figure does not propose specific methods for performing the compensation processes, but stresses the bi-directional processing, the parallel exploration of multiple alternatives, and the use of pictorial representation in the recognition process.

The flow of the computation is diagrammed more schematically in figure 10.3. The basic operation in this scheme is to seek a sequence of processing steps linking a pattern of activity (S in 10.3) in one cortical area with stored representations (such as M_1, M_2) in another. The pattern S may arise from the image of

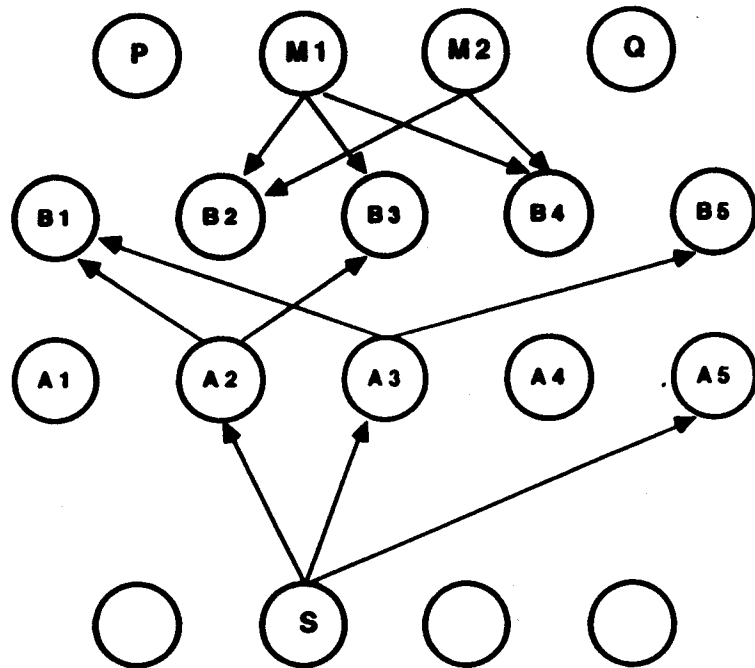


Figure 10.3

The sequence-seeking computation seeks a sequence of pattern-activations linking a source pattern (S) in one area with stored representations (M_1, M_2) in another. Nodes represent patterns of activity (co-active populations of neurons), arrows indicate how patterns activate subsequent patterns. In expanding sequences only a subset of patterns will be activated initially, and will later decay and be replaced by others. The processing is bi-directional, and a linking sequence is successfully established when the two searches meet somewhere in a large network of interconnected patterns.

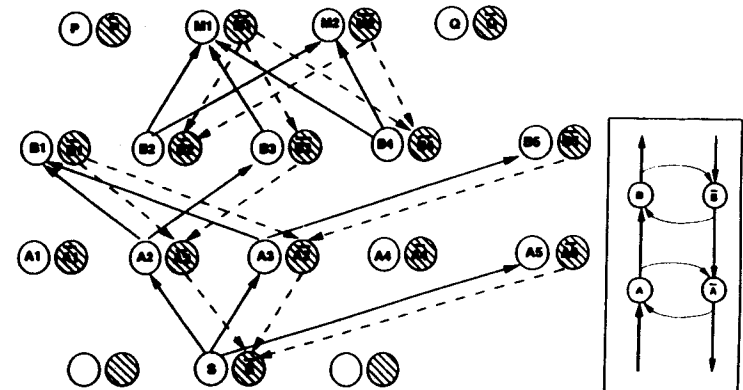


Figure 10.4

Similar to the previous figure, except that each node is split into two complementary ones. The ascending and descending streams proceed along complementary pathways. When a track is being traversed in one stream, it leaves behind a primed trace in the complementary stream. Inset shows the basic unit of the counter-streams structure. Patterns A, B on the ascending, \bar{A}, \bar{B} on the descending path. Thin arrows denote connections of the priming type. This repeating unit is embedded in a network of richly interconnected patterns.

an object, for example, a familiar face, and the patterns M_i represent stored object models, perhaps in visual area IT. As discussed above, the recognition of the viewed object involves multiple processing stages, applied in part to the incoming pattern and in part to the stored models, in an attempt to establish a match between the incoming pattern and a stored model. Intermediate patterns in the diagram correspond to different representations of the object, for example, at different 3-D orientations and scales. The figure shows a part of a network that in reality will be much larger.

Biologically, the nodes in this schematic figure represent patterns of activity of sub-populations of neurons acting together, possibly with some degree of synchrony (Abeles 1991, Engel *et al.* 1992). That is, each of the patterns, such as A_1, A_2, A_3 , is a population of perhaps a few hundred co-active neurons. As in many neural models (Hopfield 1982, Marr 1970, Willshaw, Buneman & Longuet-Higgins 1969), the populations are overlapping, that is, a given neuron can be used in more than a single pattern. The arrows in the diagram indicate how patterns activate subsequent patterns, for instance, S can activate A_2, A_3 , and A_5 . Since different patterns may share neurons, implementation constraints will place some limitations on the co-activation of patterns; for example, patterns (B_2, B_3, B_4) may be prohibited from being all active together. In expanding the sequences down from M_1 , only a subset of these patterns will therefore be activated initially, and will later decay and be replaced by other patterns.

The search is bi-directional, and a linking sequence is successfully established when the two streams of activation meet somewhere in this large network of interconnected patterns. For example, the recognition of a face image can be obtained in this scheme if processing the input image and processing the stored model can both lead to the same, or a sufficiently similar, intermediate face representation.

How can a successful link of patterns between the input and a stored model be found by the system? The proposed scheme

(figure 10.4) has two main components. First, the ascending and descending streams proceed along separate, complementary pathways. Second, when a track is being traversed in one stream, it is assumed to leave behind a primed trace in the complementary stream, making it more readily excitable, as explained further below. The scheme shown schematically in 10.4 is similar to 10.3, except that each node is now split into two complementary nodes (populations of neurons), for instance, B_2 in 10.3 is now split into B_2 on the ascending pathway and its complementary pattern \bar{B}_2 on the descending one.

The full bi-directional search now proceeds as follows. A number of sequences originating at S begin to be activated along the ascending pathway. At the same time, sequences originating at M_1 and M_2 begin to expand downwards along the descending pathway. (We will see below how some models, such as M_1, M_2 , can be selected from a larger population of stored patterns.) Whenever a track (sub-sequence) is being traversed on either stream, the complementary track remains in a primed state, ready to be activated. Not all of the possible sequences are expanded simultaneously, and already-primed patterns are activated with priority. The result will be a mechanism that searches for linking sequences in the network. Suppose that by the time S has activated A_2 along the ascending stream, the track $\bar{M}_1 \rightarrow \bar{B}_3 \rightarrow \bar{A}_2$ had already been traversed in the descending stream. This is an example of a linking taking place between the two streams: that is, a node, or group of neurons, on the ascending streams (A_2 in the example) and its counterpart (\bar{A}_2) on the descending stream, have both been activated within a limited time interval (up to a few hundred milliseconds in the case of typical recognition).

The activity will then proceed along the primed traces, since, for example, A_2 will next activate B_3 (which is primed) rather than alternative, non-primed nodes. This will result, therefore, in the immediate activation of the complete sequences $S \rightarrow M_1$ and $M_1 \rightarrow S$, establishing a complete link between the source and target patterns. This will also select M_1 as the stored pattern

corresponding to the input image S , thereby serving to recognize S as an instance of M_1 . In this manner, as a result of the priming, the top-down processing guides and paves the way for the bottom-up processing. The task of relating the sensory input to the appropriate stored representation is achieved by the cooperation of bottom-up and top-down processing, and by exploring, in each direction, multiple alternatives.

This scheme of processing is motivated directly by the proposed roles of the two processing streams, namely, that the ascending stream supports top-down, and the descending stream bottom-up processing. This proposal implies a certain degree of separation between the pathways. The separation between the streams is necessary to avoid possible intermixing between data supported by the input and states explored internally by the system. This distinction is crucial in a system that uses both top-down and bottom-up processes. For example, top-down processing can initiate the activation of an internal model (such as \bar{M}_1 in 10.4), but this event must be distinguished in the system from the activation of an internal model (M_1 in 10.4) on the ascending stream by a sequence originating at the sensory input. In other words, the bottom-up activation actually indicates the presence of the corresponding pattern in the image, whereas the top-down activation is akin to a hypothesis explored internally by the system. The separation between the processing streams is not entirely symmetric, in the sense that the descending pathway must be prevented from directly activating the ascending one, but in the opposite direction, activation of descending sequences by ascending patterns is not precluded.

The priming interaction is motivated by the assumption that recognition requires the activation of an object model stored in a high-level visual area. The matching itself can take place at any intermediate level in the network, and this match is therefore required to guide the activation towards the appropriate stored model. This is obtained in the model by the priming interaction.

The linking process described above has two additional benefits. First, a link between the ascending and descending streams can take place at any intermediate level. This has the advantage that the overall task can be split in a flexible manner between bottom-up and top-down processing. In some cases, such as the recognition of a highly familiar object, the process will be primarily bottom-up, since compensation for viewing direction, illumination, and the like, will not be required. Less familiar views will require a more substantial contribution of the top-down processes, and the relative contribution of the two processes can change in the system from one situation to another. Second, to establish a link, the ascending and descending patterns need not arrive at a given node simultaneously; a meeting is also possible between an active pattern and a pattern that had been active some time before and decayed, but left a primed trace in the complementary stream. This is convenient because strict coincidence of activated patterns is not required.

In terms of connectivity, the excitatory connections between patterns are predominantly reciprocal, obeying the following general rule (figure 10.4b): whenever A is connected to B , there is a back-connection from \bar{B} to \bar{A} , with cross-connections between A and \bar{A} and B and \bar{B} . (Inhibitory connections also play a role, but will not be discussed.) The cross-connections are assumed to have a priming effect: when B , for instance, is activated, it also provides input to \bar{B} , making it more readily excitable by a subsequent input along the descending stream. The reciprocity of the connections is an inherent aspect of the model, and it is also a distinguishing feature of cortical connectivity (although some exceptions have been noted, Distler *et al.* 1991, Rockland, Saleem & Tanaka, 1992). It should also be noted that although the counterstreams structure uses "forward" and "backward" connections, it does not necessarily imply a simple hierarchical structure; it can incorporate a more general structure as long as the above connectivity rule is obeyed.

In summary, the model proposes a general form of computation, called sequence-seeking, and a particular structure, called counter-streams, that supports the required flow of information. The sequence-seeking process has two main characteristics: it is bi-directional, and it explores multiple alternatives simultaneously. The basic structure of the counter-streams model is relatively straightforward, comprising two complementary networks going in opposite directions, with interaction between them primarily (but not exclusively) in the form of enhancing patterns across the two streams. In later sections of this chapter we will examine in more detail the proposed relation of this structure to cortical circuitry.

10.1.4 The Role of the Anatomical Back Projections

A point worth stressing is the role assigned by the model to the descending, or feedback, projections in the visual cortex. These connections have attracted considerable attention since it was found that they tend to be as massive as the forward pathways, and they tend as a rule to reciprocate the forward connections (that is, when a forward connection exists between two cortical regions, the reciprocal back projection almost always exists as well). This pattern of connections is surprising if one considers visual processing as an essentially bottom-up process.

The current model makes a simple proposal: the top-down pathways are used as the anatomical substrate for top-down processing. This role can be contrasted with other models, where the descending projections are used for different purposes: controlling selective attention (Fukushima 1986, Koch 1987), grouping and figure-ground segregation (Okajima 1991, Sporns, Tononi & Edelman 1991), learning processes (Zipser & Rumelhart 1990), modulating cortical output to other visual centers (Sandell & Schiller 1982), or to correlate and synchronize the activity of interrelated neuronal groups (Tononi, Sporns & Edelman 1992).

In proposing this role for the descending pathways it is worth noting that the use of top-down processing in the proposed model

is more extensive than in alternative models of visual processing. As already mentioned, top-down processing is sometimes taken to refer to the use of high-level contextual knowledge in perceptual processing. For example, in an office scene one expects to find certain objects such as a desk, a chair, and a telephone, and top-down processing refers to the use of such expectations in the perceptual process. In this view top-down processing is more cognitive than visual in nature. The visual processing is performed mainly by bottom-up processing applied to the incoming image, and top-down processing becomes effective primarily when the visual input becomes ambiguous, as a result of poor lighting conditions, severe occlusion, and the like. In contrast with this limited role, computational studies of visual recognition have shown the need and the feasibility of using top-down processing as an integral part of the recognition process. The use of top-down processing in this view includes the use of stored information required to deal with the effects of viewing direction, illumination, occlusion, and object deformation. Consequently, the top-down processing is expected to be as extensive as, and roughly symmetric to, the bottom-up part.

10.1.5 Model Selection

The bi-directional process raises an important question regarding the activation of stored models for top-down processing. To initiate appropriate top-down processing, some initial selection and subsequent refinement of a relevant subset of stored models is required. The next two sections discuss two mechanisms for this task, both supported by evidence regarding human perception.

Initial Classification One mechanism of model selection is provided by the mechanism of initial classification. It has been suggested in chapter 6 that as an intermediate stage on the way to individual identification an object is often classified first more broadly as a face, a car, a bird, and the like. The classification may even be non-unique, that is, a number of competing interpre-

tations may still exist at this stage. Following the initial classification some stored models will become more likely than others, and will be activated and processed with higher priority. For example, an object may be classified as a face prior to its individual identification, and following classification, face-related sequences will be expanded preferentially. Biologically, fast classification will involve the activation (or inhibition) of high-level patterns on the descending stream by low-level ascending patterns, resulting in the preferred activation of the selected patterns. This interaction is similar to the priming interaction between the pathways, with the exception that in this case the ascending pathway can activate, rather than facilitate, the descending one. The fast initial selection of subsets of models will not be limited to the activation of object models at a single "topmost" level; intermediate models at different levels along the descending stream can also be activated and serve as the starting points for descending sub-sequences. For example, in addition to the selection of a complete face model, intermediate models of face-parts can also be activated, and perhaps also stored models of some basic image configurations of the type described by Fujita *et al.* (1992) in the study of visual area IT. To achieve fast initial selection, this process may use some of the direct connections that are known anatomically to take a relatively short route on the way from low to high visual areas (such as the connections from area V4 to AIT, or from V3 and VP to area TF; Felleman & Van Essen 1991).

The Use of Context A second mechanism for model selection is provided by the effects of expectation and context. The essential idea is that temporal and spatial correlations can influence the likelihood of different models. Knowledge about the current situation can thereby be used to influence the activation or priming of a subset of models that will then become preferential sources for descending sequences.

As discussed at the end of chapter 6, context can have a powerful influence on the processing of visual information (as well

as in other perceptual and cognitive domains). A pair of similar elongated blobs in the image may be ambiguous, but in the appropriate context, for instance, under the bed, they may be immediately recognized as a pair of slippers.

Familiar objects can often be recognized in the absence of context, but in dealing with less familiar objects, or with complex scenes, or when the viewing conditions are degraded, the role of context increases in importance and can become indispensable. Even when context is not strictly required, the appropriate visual context still facilitates the recognition process, and makes it faster and more reliable (Biederman *et al.* 1982, Palmer 1975, Potter 1975). Context information that helps the observer expect a certain class of objects facilitates recognition significantly, and when objects are placed in an unusual context, recognition is hampered. Under natural conditions, useful context information is almost always present, and this accounts in part for our capacity to deal effectively with complex scenes.

Context effects can operate in the framework of the sequence-seeking scheme by the prior priming of some of the patterns (populations of neurons). The effect will be similar to the mutual priming of the ascending and descending streams, but over longer time-scales. (Priming between the streams may last for tens to hundreds of milliseconds, context effects should last for considerably longer, up to minutes or hours.) Sequences passing through the primed patterns will then become facilitated. In the above example, the location of the blobs, under the bed, will prime patterns representing objects that are commonly found in that location, making slippers a likely interpretation.

The general notion of priming internal representations is a common one (Kosslyn 1994), but its effects in the framework of the sequence-seeking scheme are particularly broad. When certain patterns are activated, for instance, by noticing and identifying the bed in the image, they will initiate sequences of their own, and an entire set of patterns will end up in a primed state. Later on, other sequences passing through a primed trace will be facil-

itated, compared with the non-primed sequences. The resulting effect is that a context pattern *A* may help to bring about the activation of *B* not as a result of direct pre-wired association, but because an intermediate sub-sequence leading from *A* to *B* had been previously facilitated. Context effects will therefore have indirect and wide-spread influence.

The spread of context effects may capture some of the fundamental aspects of context effects in humans. Human perception and cognition appear to have an almost uncanny capacity (which is remarkably difficult to reproduce in artificial systems) for bringing in relevant context information in a broad and flexible manner. It seems that broad and indirect context effects of this kind can be reproduced by the sequence-seeking computation.

10.1.6 Learning Sequences

Recognition in the sequence-seeking scheme can become faster and more efficient by the learning of past successful sequences. A successful sequence is a sequence of pattern activations linking an input pattern with a stored model. When faced again with a similar input, the computation will follow the sequence that proved successful in the past rather than search anew for a possible link between the input and a stored representation.

The counter-streams structure makes it possible to use a simple and local learning rule to reinforce selectively complete, successful sequences. The reason is that every pattern along a successful sequence will receive both a direct activation and a priming signal from the complementary track. In contrast, patterns on dead-end tracks will receive one or the other, but not both. The approximate temporal coincidence of the two signals can therefore be used to preferentially strengthen the successful sequence. This rule is local, since it depends on the activation of a single pattern. Yet it is sufficient to reinforce preferentially successful sequences forming an uninterrupted link between source and target patterns. Following practice, out of the huge number of possible sequences,

those that proved useful in the past will be explored with higher priority in future uses of the network.

From a biological standpoint, this is a favorable and perhaps surprising property of the counter-streams structure. As in other models of learning in neuronal networks, learning is accomplished by the modification of synaptic efficacy, and the change is determined locally by the activity of the pre- and post-synaptic units. One might expect difficulties in using such a local rule to distinguish complete from dead-end sequences, since this involves a global distinction that depends on the entire sequence of activation. Yet, because the counter-streams structure combines patterns of activation flowing in both directions, the local rule is sufficient for learning globally successful sequences. The synaptic mechanism responsible for the learning of sequences is also expected to have some special properties not found in more standard Hebbian models. In particular, the synaptic modification is expected to be more effective in neurons in the primed compared with the non-primed state.

In the process of reinforcing successful sequences, changes due to learning are distributed throughout the system, and are not confined to high-level centers specializing in learning (Sejnowski 1986). Recent studies of learning certain perceptual skills suggest that low-level visual areas are indeed involved in the modifications that take place during the learning process (Karni & Sagi 1991).

In addition to the learning of complete sequences, as above, the system may also be engaged in the learning of the individual stages, that is, the different steps comprising the processing sequences. This aspect of the learning is treated, for example, by Poggio (1990). However it remains outside the scope of the current discussion, since the focus here is not on the specifics of individual processes, but on their overall common structure.

10.1.7 Searching for the Best Sequence

Due to the parallel exploration of multiple alternatives, and to the tuning of the system by past experience, straightforward recogni-

tion tasks will require little or no search. More complex tasks will require a search through the space of possible sequences for the appropriate transformations that will bring the viewed object and the stored model into close alignment.

In previous chapters we have discussed possible methods for determining the required transformations between the stored model and a novel view. For example, one possible method for dealing with the effect of viewing direction is to determine the required transformation uniquely on the basis of a small number of corresponding features. Another possibility is simply to try out a number of alternatives and then select the most appropriate one. Consider for example the problem of compensating for a possible scale difference between a novel view and a stored object model. This can be accomplished by first recovering the required transformation uniquely based, for example, on the matching of corresponding features, and then applying the normalizing transformation to the viewed object or the stored model. An alternative approach, mentioned earlier in this chapter, is to perform a multi-scale analysis by generating internally a number of copies at a number of different scales, and then selecting the best-matching one. To avoid excessive search, it will be useful to first perform a rough alignment, as discussed in chapter 7, that will narrow down the range of scale corrections that must be applied. Similarly, in compensating for 3-D viewing direction, illumination, and object deformation, instead of recovering uniquely the required transformations, the alternative approach is to generate and test in parallel a number of competing alternatives. As discussed in previous chapters, this will be based on the use of stored information regarding possible object changes, associated with both specific objects and object classes. This simple search approach replaces a sequential and relatively sophisticated computation by a simpler, parallel, but more extensive process. For standard computers, the sequential sophisticated computation is usually the method of choice. For a biological system, however, the simpler parallel approach is probably more suitable.

We have tested in simulations some aspects of a process that searches in parallel for the appropriate alignment transformations. The overall structure of the search is simple. It tries out in parallel a set of N competing possibilities. The size of this set will be determined in practice by hardware limitations. If none of the alternatives produce a sufficiently close match, the process selects a subset of k patterns, and explores a new set of N patterns around the selected subset. The process uses the fact that the matching between patterns is not an all-or-nothing event, but a graded one. Some sequences will lead to better matches than others, and will then serve as starting points for exploring additional sequences, leading in turn to an improved match. This process is related to the method of Bayesian optimization (Mockus 1989), and also has some features in common with a family of optimization and search procedures known as "genetic algorithms" (Holland 1975, Goldberg 1989). Recent evaluations have shown such methods to behave quite efficiently (Brady 1985, Peterson 1990). Our own simulations in the context of pattern matching have also shown that computations based on sequence-seeking compare favorably with alternative methods, such as gradient descent and simulated annealing.

Some of the simulations were applied to a database of Japanese characters shown in figure 10.5. The search in this case was in a six-dimensional transformation space, because the input characters could be scaled, rotated, shifted, stretched, and sheared with respect to the stored patterns.

Figure 10.6 shows a simulation of a simplified sequence-seeking process applied to such 2-D patterns. This simulation, as well as the other examples in this section, were performed by A. Zeira at the Weizmann Institute. The example is intended not as a realistic model, but to illustrate the process in a simple example. The task is to recognize an input shape (example in 10.6a) by comparing it with stored shapes. To make the task more difficult, noise was added to the input images. The search in this example was two-dimensional—the input shape was displaced in the x ,



Figure 10.5
A database of characters used in the simulations of the sequence-seeking process.

y directions with respect to the stored pattern. The objective function to be minimized by the search, which is the degree of match as a function of displacement, is a complex function that contains multiple minima. Level-contours of this function (that is, contours of constant function values, similar to elevation contours) are shown in 10.6b. The correct solution is represented by the deep valley in the middle, but there are many additional local minima.

To determine the optimal match between the input shape and the stored pattern, one possibility is to displace the input shape by different amounts in x and y , and compare each displaced version with the model until the best match is obtained. Instead, the simulation used a different strategy to test aspects of the bi-directional search employed by the sequence-seeking computation. The search in the x and y directions was divided into two parts. The model M was shifted in x , to generate a number of copies M_i at different horizontal locations. The input image was shifted in y , generating displaced copies I_k . This is a simplified example of splitting the computation between the two directions. In more realistic cases of 3-D recognition the internal transformations applied to the model will be more complex, and will depend, for example, on its 3-D structure. The search proceeds by comparing the displaced versions of the input and stored patterns. A good match between a pair M_i, I_k , then leads to the generation of new copies around the corresponding displacement x_i, y_k . The results were compared with the one-directional version, where M remained fixed, and copies of the image $I_{j,k}$ were generated by shifting I in both the x and y -directions.

The procedure used to generate the new “offsprings” around the existing patterns was a simple genetic-like algorithm: the likelihood of generating a new sample increases near good past samples (as determined by the function f_1 in 10.6c), and decreases with the density of past samples. (The shape of f_1 was guided by theoretical considerations that will not be detailed here.) This function is computed in the vicinity of past samples. For example, if a good match was obtained between the input displaced horizon-

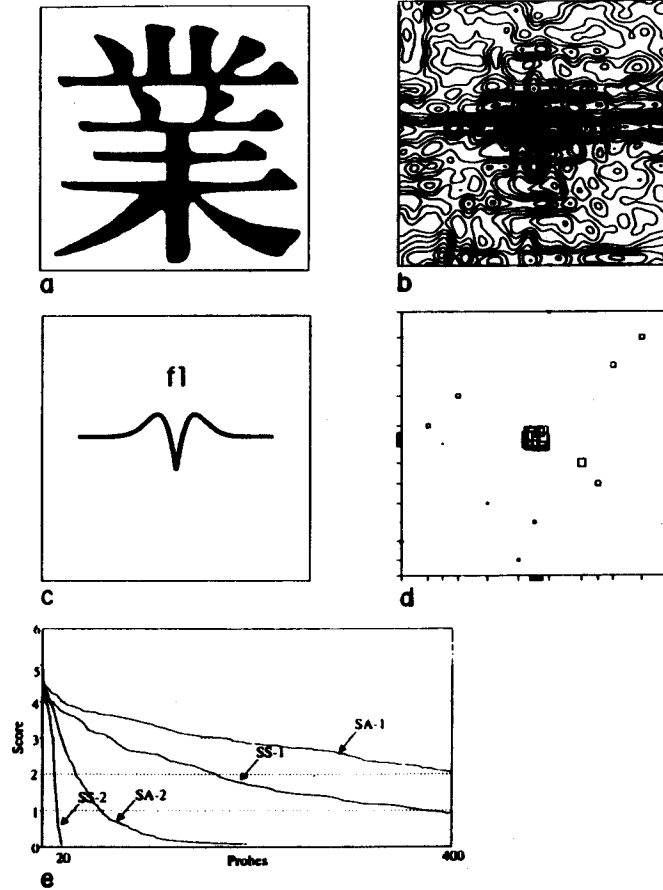


Figure 10.6

Simulation of a simplified bi-directional process, used to match a noisy input shape by comparing it with stored shapes (example in a). The input shape is displaced in the x, y directions with respect to the stored pattern. The objective function to be minimized is complex and contains multiple minima (level contours in b). The likelihood of generating a new sample increases near good past samples, and decreases with the density of past samples. Bias towards good sample points was determined by f_1 in c. (d) A representation of the search following 20 displacements in x and y . The marks show the selected displacements, the squares show the degree of match obtained at some x, y displacements, coded by size (For each x_i , the best matching y_k is shown, similarly for the y_j 's). The search combines exploration of the domain with concentration around good matches. (e) Match quality as a function of number of patterns explored for one (SS-1) and two-directional (SS-2) versions of the sequence seeking search, compared with one- (SA-1) and two-directional (SA-2) versions of simulated annealing. Reproduced from Ullman (1995).

tally by \bar{x} units and the stored pattern displaced vertically by \bar{y} units, new displacements will be generated around the successful values (\bar{x}, \bar{y}) . However, if the match was poor, or if many patterns with a similar displacement were already compared, then the likelihood of attempting further solutions in the vicinity of (\bar{x}, \bar{y}) will decrease. In this manner the past samples induce over the search-space a likelihood function, and the next samples occur at maxima of this function. The process is simple: it proceeds by trying a number of alternatives, and then selecting and refining successful solutions.

The simulations of the simplified bi-directional search show that the process has a number of favorable general properties. First, the search locates the optimal match efficiently, as shown in 10.6e. (The score is in units of σ , the standard deviation of the terrain in b.) It proved more efficient in the pattern matching task than commonly used minimization methods such as simulated annealing (SA in 10.6e) or gradient descent using multiple starting points. Second, the bi-directional scheme in these examples is considerably more efficient in terms of the number of patterns explored than a one-directional process (SS-1 vs. SS-2 in 10.6e). This advantage will hold as long as the number of stored patterns to be explored is not too large. Third, the use of past results in guiding the search biases the process to concentrate in more promising regions, compared with SA and gradient descent. Finally, this process exhibits good capacity to escape local minima in reaching for the global solution.

The examples used a simplified task, but the search-space was of significant size (up to a million locations) and contained multiple local minima. Further experiments were also applied to face images, as shown in figure 10.7. In this case the stored model consisted of two face images, shown at the top of the figure, with correspondence established between them. The transformations explored by the process included rotation in depth around the vertical y -axis, scaling, translation, and rotation in the image plane.

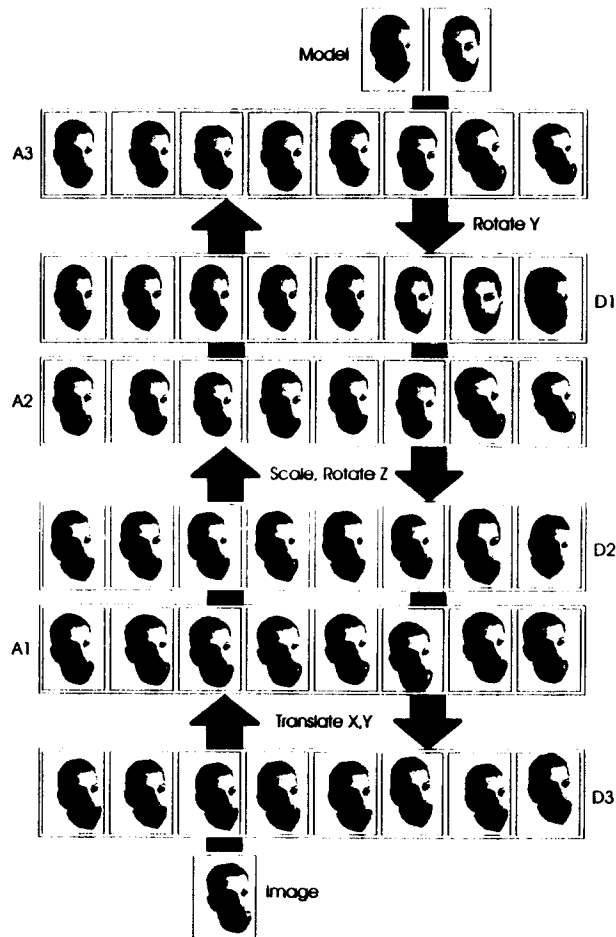


Figure 10.7

The sequence-seeking process applied to face images. The stored model consists of the two top images, with known correspondence. The transformations performed during recognition included rotation in depth around the vertical y -axis, scaling, translation, and rotation in the image plane. A1, A2, A3 are patterns on the ascending stream, D1, D2, D3 on the descending one.

The rotation in depth was obtained from the model images using the view-combination method. In the ascending direction, rotation in depth was not applied. The use of such transformations, that are applied in one direction only, requires some additions to the scheme that will not be detailed here. As before, the scheme performs the recognition task without establishing feature correspondence, and without explicitly recovering the transformation parameters. The scheme simply tries multiple alternatives, and selects the best ones.

The computational experiments suggest that a search of this type, combining bi-directional search with the exploration of multiple alternatives can be quite efficient. At the same time, it is worth stressing that in many recognition tasks the search is not expected to be extensive, for several reasons. First, the use of rough alignment will limit, as mentioned above, the range of the required compensating transformation. Second, the use of past experience can lead to a direct and immediate match. Suppose that V is a view of a familiar, frequently seen object. The view V itself may then become a part of the stored object representation, eliminating the need for compensating transformations. Alternatively, the sequence required to match V with the appropriate stored model can be learned and reinforced with time. This sequence will be consequently explored with high priority, eliminating again the need for search. In the model discussed above, the acquired increased efficiency with practice will not be limited to the view V itself, but will generalize to similar views: if the input view is similar to V , the search will still be efficient because it will use the learned sequence as a starting point and explore additional sequences in its vicinity. In conclusion, the compensating transformations required to match the input with a stored model will usually require little or no search. When a search becomes necessary, it will be performed efficiently by the sequence-seeking process.

10.1.8 Generic Aspects of Sequence-Seeking

The discussion of the sequence-seeking process focused on the domain of visual recognition. However, the process of establishing a sequence of transformations, mappings, or states, linking source and target representations, provides a useful general mechanism for various aspects of perception as well as for non-perceptual functions. For example, the planning of a motor action can be cast at some level in terms of seeking a sequence of possible moves linking an initial configuration with a desired final state. Movement trajectories could be based in a sequence-seeking scheme on a repertoire of elementary movements, and these basic movements will then be transformed (scaled, stretched, rotated, etc.) and concatenated together to generate more complex movements. In analogy with sequence-seeking in vision, movement planning could also utilize a bi-directional search that explores in parallel multiple alternatives. Similarly, more general planning and problem solving can also be formulated in terms of establishing a sequence of transformations, mappings, or intermediate states, linking some source and target representations (as proposed, for example, by Newell & Simon's (1972) GPS model, Quillian's (1968) semantic net theory, see also Winston 1992), and they may therefore benefit from computations similar to the sequence-seeking scheme. I will not discuss these general problems further, beyond raising the possibility that general aspects of the sequence-seeking process provide a useful computational scheme that could be applied, with appropriate modifications, to different cognitive tasks. This possibility is consonant with the widespread hypothesis (e.g., Barlow 1985, Creutzfeldt 1978, Crick & Asanuma 1986, Edelman 1978, Martin 1988a, White 1989) regarding the possible existence of some general cortical mechanisms that are applicable, with suitable local modifications, to a broad range of different tasks.

10.2 Biological Embodiment

In this section, biological aspects of the model will be discussed. It is suggested that the general connectivity structure of the cortex is highly suitable for supporting the bi-directional, multi-path computation of the sequence-seeking model. The model is used to interpret key aspects of cortical connectivity, to derive new predictions, and to raise problems for further study. The discussion will focus primarily on general features of the model, such as the ascending and descending streams, and the laminar distribution and general pattern of connections between and within cortical areas. Several aspects of a more specific nature will also be considered, to illustrate possible predictions and questions for further study.

The sequence-seeking model requires two pathways going in opposite directions with the appropriate cross-connections. A schematic diagram proposing how the counter-streams structure may be embedded in cortical connections is shown in figure 10.8a. The proposed embodiment is presented in schematic outline only, focusing on a number of central aspects, but without discussing details or possible variations of the model.

The ascending stream goes through layer 4, which is the main input layer in the cortex, to a sub-population of the superficial layers above it, denoted in the figure as AS (for Ascending Superficial), and then projects to layer 4 of the next cortical area (II in the figure). The descending stream goes through a different sub-population of the superficial layers (DS, for Descending Superficial) to DI (for Descending Infra), a subpopulation of the infragranular layers (often in layer 6), and from there to DS of a preceding area. The connections can also skip one step (or occasionally more) in the stream, such as AS directly to AS on the ascending stream, and DS → DS or DI → DI on the descending stream (thin lines in figure 10.8a.)

Layer 5 is left out of the diagram because, according to the model, this layer (or a part of it) is involved primarily not in the

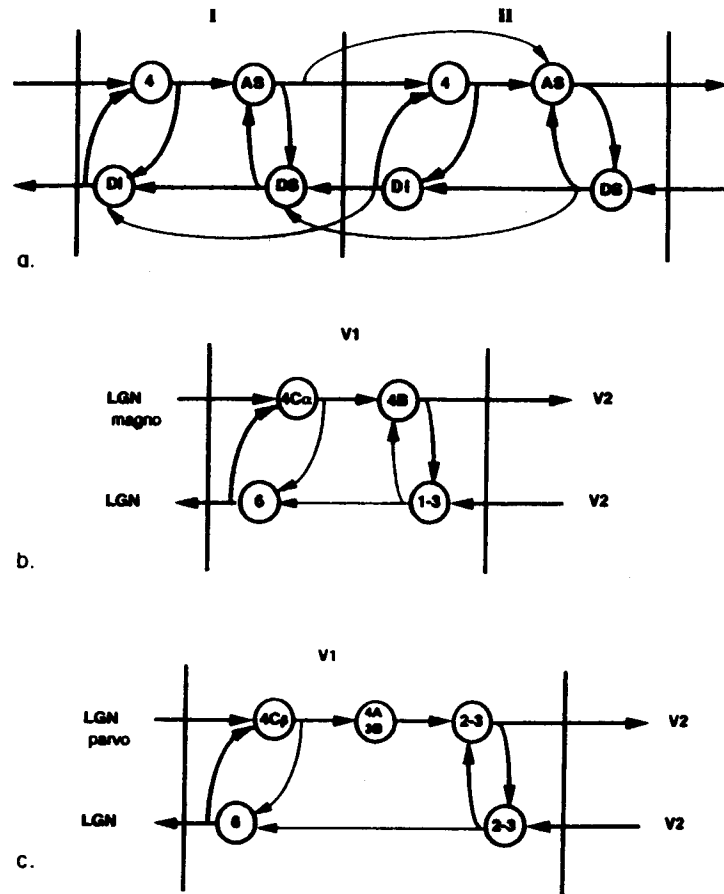


Figure 10.8

a. How the basic counter-stream structure may be embodied in cortical connectivity. The structure contains two interconnected streams, an ascending and a descending one. The ascending path goes through layer 4 and the ascending superficial population (AS) to the next area. The descending path goes from the descending superficial (DS) population to DI (descending infra) and back to the first area. Thin arrows show pathways that "leap over" a step in the stream. Inhibitory and long-range intra-areal connections are not shown. See text for more details. b. A schematic representation of the main connections according to the model along the magnocellular stream from the LGN via V1 to V2. (V1 is also connected to other visual areas, not shown in the diagram) The connections are drawn in a manner suggested by the model and a above. Thick arrows: established connections; thin arrows: connections predicted by the model. c. The main connections according to the model along the parvocellular stream from the LGN via V1 to V2. Thick arrows: established connections; thin arrows: connections predicted by the model.

main streams, but with their control, in cooperation with sub-cortical structures. There are at least two reasons for assuming that layer 5 (or parts of it, such as 5b of the macaque's V1) may be involved in control functions. First, its orderly connections to subcortical structures (such as from visual cortex to the pulvinar and the superior colliculus, structures implicated in controlling attention and eye movements, Desimone *et al.* 1990) that are reciprocally connected in turn in a topographic manner to multiple visual areas. Second, the firing pattern of a population of pyramidal cells in this layer that "...can initiate synchronized rhythms and project them on neurons in all layers" (Silva, Amitai & Connors, p. 434).

Note that the counter-streams structure suggests a natural organization in about five to six main layers: one or two performing control functions, two (an input and an output layer) for the ascending and two for the descending streams. Furthermore, the main layers are assigned in the model a characteristic functional role, and this appears to be compatible with the relatively uniform pattern of the laminar distribution of inter-area connections. The division between the roles of the different layers is likely to be in reality less clear-cut, and there are known variations and specialized sub-laminations. However, the goal of the diagram is to emphasize the possible common underlying structure according to the model, rather than to account for possible variations.

It is interesting to note that from a developmental standpoint the layered cortical structure appears to develop in two stages, possibly from distinct origins (Marin-Padilla 1978, Deacon 1990). The most superficial and deepest layers develop first, and all other layers develop subsequently in between them. These two structures may be the precursors of the descending and ascending streams, respectively (Deacon 1990, Mumford 1993). This developmental view is compatible with the notion of the two distinct and interconnected streams constituting the two main building blocks of the counter-streams structure.

10.2.1 Connections of V1: Data and Predictions

To give a more specific example, figure 10.8*b, c* shows an expanded version of the connectivity diagram, applied to cortical area V1 (which is somewhat special, but for which the data are more comprehensive than for other visual areas), and its connections to the LGN below it and cortical area V2 above. V1 is also connected to other visual areas that are not shown in the diagram. The input to area V1 is comprised of two sub-populations of cells in the LGN, the magno- and parvocellular inputs.

Figure 10.8*b* shows the connections in the macaque of the magnocellular stream, 10.8*c* of the parvocellular stream, (Rockland and Lund 1983, Lund 1988*a, b*, Martin 1988*a*). The diagram shows the main connections; some additional secondary ones exist but will not be considered. The connections are drawn in a manner suggested by the model, and they include both known connections (thick arrows) and connections predicted by the proposed scheme but for which empirical evidence is partial or lacking (thin arrows). As can be seen, the pattern of connections in the two streams is in general agreement with the counter-streams structure and figure 10.8*a*.

If the general hypothesis regarding the counter-streams structure is broadly correct, then a number of predictions can be made regarding the main connectivity patterns within and between areas. One general prediction is the possible distinction between the AS and DS sub-populations. This separation reflects the most straightforward implementation of the scheme; however, some alternatives can exist without violating the constraints of the model.

A separation between the ascending and descending populations is evident in the connections involving layer 4: the ascending projections terminate in layer 4, the descending projections always avoid it. In the superficial layers the situation is more difficult to assess, and the available evidence is at present restricted. In the magnocellular projection from V1 to V2, this separation is respected—the forward projection originates mainly in 4B, while the

back projection is mainly to other layers (figure 10.8*b*). It is further expected that even when the superficial layers provide both the source and the target of connections to another area, there will in fact often be a separation to the AS/DS sub-populations. If these populations exist, they should be connected in a reciprocal manner. A related expectation derived from the model is the existence of priming-typesynaptic interactions. That is, excitatory synaptic input that by itself may not be very effective in driving the target cells, but that facilitates the effects of subsequent inputs to these cells.

An example at the other end of the spectrum, that is, a highly specific expectation, is that in the magnocellular stream the model suggests reciprocal interconnections between layer 4B (playing the part of AS in the model), and layers 1-3, the recipients of descending projections from V2 (DS in the model). Moreover, the same superficial cells connected to 4B will also be the recipients of descending projections from V2. The projection from 4B to the superficial layers is well established. It is also known (Lund 1988*a*) that 4B pyramidal cells send apical dendrites to the superficial layers where the connection may take place.

The model also includes a reciprocal connection between layer 4 and the LGN-projecting cells in layer 6. The projection from 6 to 4 is well-established in both the cat (McGuire *et al.* 1984) and monkey (Lund 1988*a*), and there is support for the opposite connection as well (Lund & Boothe 1975). It is also interesting to note in this regard that the population of layer 6 cells projecting back to the LGN were found (in the cat) to be the same cells that are also connected to layer 4C, by axonal collaterals and dendritic arbors (Katz, Burkhalter & Dreyer 1984), in accordance with the connectivity in 10.8*b, c*.

The connections between layers 4 and 6 are expected to have a priming effect in the model (not necessarily the only effect, see Bolz & Gilbert 1986, Martin 1988*b*), and this notion has some physiological support. It was found (Ferster and Lindström 1985) that using electrical activation of layer 6 cells by antidromic ac-

tivation increased the probability of layer 4 firing, and most cells fired multiple spikes in response to each ascending stimulation. Under the opposite conditions, when layer 6 was inactivated, the main observed effect was the reduction in excitability of layer 4 cells (Grieve, Murphy & Sillito, 1991). This priming of layer 4, which is the main input layer to the visual cortex, can be viewed as reflecting the expectations of the visual cortex, attempting to detect expected patterns in the image.

From an anatomical standpoint, EM reconstructions (McGuire *et al.* 1984) have shown terminations of layer 6 axons on smooth and sparsely spiny cells. In the cortex, smooth cells are usually inhibitory and spiny cells are usually excitatory. These findings were therefore interpreted as connections to inhibitory interneurons in layer 4, unlike the prediction of the model that suggests a projection onto layer 4 spiny cells. Detailed recent data by Ahmed *et al.* (1994) support, however, the model's prediction, and indicate that the major target of layer 6 pyramidal output to layer 4 are in fact the spiny stellate cells rather than inhibitory interneurons.

Layer 6 is also involved in the model in the descending pathway (although, as mentioned above, layers 5 and 6 are often further divided into distinct sublayers, and the identification of layer 6 with the Descending Infra population in the model is not always straightforward.) The involvement of layer 6 in the descending pathway, as either the origin or the target of the descending projection, has been demonstrated in many visual areas, including V1, V2, V3, V4, TEO, MT, MST, VIP, PO, LIP (Andersen *et al.* 1990, Felleman & Van Essen 1984, Maunsell & Van Essen 1983, Lund *et al.* 1981, Rockland, Saleem & Tanaka 1992, Rockland & Virga 1989).

10.2.2 Lateral Connections between Areas

Connections between cortical areas (not only visual, but also somatosensory and motor) can be classified into "forward," "backward," and "lateral" connections, on the basis of the laminar distribution of their source and destination (Rockland & Pandya

1979, Maunsell & Van Essen 1983, Friedman 1983, Van Essen 1985, Zeki & Shipp 1988, Andersen *et al.* 1990, Boussaoud *et al.* 1990, Felleman & Van Essen 1991). Lateral connections terminate in all layers, and their origin is bi-laminar, from the supra as well as infra layers. (These lateral connections between areas should not be confused with the horizontal connections within a region.) The lateral pattern is relatively complex; it is therefore interesting that a number of its main features can be derived almost directly from the model. The counter-streams structure does not require a distinct, third type of connections. It allows, however, forward and backward connections simultaneously in both directions, and it can include lateral connections by simply assuming that they are the union of ascending and descending connections. If this view is correct, then the main connections participating in the lateral connection can be inferred from the basic scheme (figure 10.8a). According to the model, they include the direct connections: AS \rightarrow 4, and DI \rightarrow DS, as well as the connections that leap over one stage in the diagram, namely, AS \rightarrow AS, DS \rightarrow DS, DI, and DI \rightarrow DI.

The origin of the projections according to the model would be bi-laminar, and the terminations would span all layers, in agreement with the observed pattern. This can also provide an explanation for the problem of irregular terminations (Felleman & Van Essen 1991), that occurs when the pattern of terminations is at odds with the usual distinctions between forward, backward, and lateral connections. This can happen, for example, when some of the terminations are restricted to layer 4 of the target area while others show columnar terminations. This pattern was termed F/C (for a mixture of "four" and "columnar") paradoxical termination, since termination in layer 4 is a signature for ascending connections, while a columnar termination signifies lateral connections. In the counter-streams structure, the point to note is that the lateral connections from the superficial layers of area *A* to target area *B* are composed of two sub-projections: ascending (AS \rightarrow 4) and descending (DS \rightarrow DS, DI). Anterograde labeling of the

upper layers of area *A* can therefore show mixed patterns of terminations, such as 4 alone, or a columnar termination, in agreement with the F/C paradoxical termination. It can also (by labeling the DS alone) show a bi-laminar pattern of connections, and this can account for the other types of irregular terminations. conclusive

The detailed nature of these connections is still not entirely clear. However, the proposed account serves to illustrate two points. First, that some of the apparent complexities may have a natural explanation within the counter-stream structure. Second, if the account is generally correct, it provides support for the existence of the AS and DS subpopulations in the model.

10.2.3 Priming Mechanisms

A central prediction of the model concerns the existence of synaptic interactions of the priming type. These are synaptic inputs that by themselves will not be sufficient under normal conditions to drive the target cells, but will increase the efficacy of subsequent inputs to these cells. A more conventional view is that if the synaptic input fails to reach threshold it will decay and die out, and will have little effect on subsequent processing. Here the expectation is that some sub-threshold interactions will have long-lasting effects on subsequent processing.

Although priming interactions have not been studied directly, some known or physiologically plausible mechanisms could play a role in such priming interactions. Priming can be obtained for example by long-lasting depolarization, combined with subsequent input, added either linearly or nonlinearly. A long-lasting depolarization can be caused by a number of possible mechanisms, including the activation or inactivation of ionic channels with a slow time course (Amitai *et al.* 1993, Hirsch & Gilbert 1991, Markram & Sackman 1994, McCormick 1990, Wilson 1995), NMDA receptors (Miller, Chapman, & Stryker 1989), or the activation of distal parts of the dendritic tree (Stratford *et al.* 1989). This depolarization will facilitate subsequent inputs by summation (Miller, Chapman, & Stryker 1989), or by a nonlinear interaction (Esguerra

Kwon, & Sur 1989, Koch 1987, Sherman *et al.* 1990). From the sequence-seeking model one might expect in fact to find not a single mechanism, but a number of different ones, operating at a range of different time-scales. Related to the priming effect, one might also expect a "reset" mechanism, that resets primed neurons to a non-primed state following activation. Although the details are not known, it appears that synaptic mechanisms for priming connections are physiologically plausible, and it will be of interest to try to test them empirically.

10.2.4 Effects of the Feedback Projection

According to the sequence-seeking scheme, the physiological effects of the descending projections can assume two different forms: either the priming and modulation of the ascending stream, or the direct activation of a lower area. Both effects have been observed in physiological studies, modulatory (Nault *et al.* 1990, Sandell & Schiller 1982), as well as direct excitatory effects (Mignard and Malpeli 1991, Cauller & Kulics 1991). They are also supported by functional imaging studies showing that low-level visual areas can be activated by tasks such as visual imagery (Kosslyn *et al.* 1993). Further predictions of the model regarding the modulatory effects include: (i) similar modulatory effects are also likely to be exerted by ascending signals on descending ones, (ii) the two effects of the back-projections may be segregated into two distinct sub-populations: in figure 10.4b, *B* can be directly driven along the descending stream, but patterns such as *B* on the ascending stream are expected to show modulatory effects.

10.3 Summary

In summary, the computation proposed by the sequence-seeking model is a bi-directional process performed by the combination of top-down and bottom-up streams of processing. Bottom-up processing is supported by the ascending pathways, top-down processing by the descending ones. In each direction, different

alternatives are explored in parallel. The scheme incorporates a number of basic lessons from computational vision and perceptual psychology. Essential properties of the scheme include the simultaneous exploration of multiple alternatives, the relatively simple, uniform, and extensible structure, the flexible use of 'bottom-up' and 'top-down' sequences that can meet at any level, the roles of context and of fast classification, and the learning of complete sequences by a simple local reinforcement rule.

The model combines the proposed computation with a number of known as well as predicted aspects of cortical circuitry. Given the still limited knowledge regarding cortical structures and the computations they perform, the model addresses mainly general aspects of the computation. The combination of the proposed computation and structure serves to suggest a framework that offers a computational account for several basic features of cortical circuitry, such as the predominantly reciprocal connectivity between cortical areas, the forward, backward and lateral connection types, the regularities in the distribution patterns of inter-area connections, the organization in 5-6 layers, and the effects of back projections, as well as a number of more specific details. It also poses problems for further study at the structural as well as computational levels.