Building a 21st century organization

Business, Organization

The power and versatility of the human visual system derive in large part from its remarkable ability to find structure and organization in the images encoded by the retinas. To discover and describe structure, the visual system uses a wide array of perceptual organization mechanisms ranging from the relatively low-level mechanisms that underlie the simplest principles of grouping and segregation, to relatively high-level mechanisms in which complex learned associations guide the discovery of structure.

The Gestalt psychologists were the first to fully appreciate the fundamental importance of perceptual organization (e. g. , see Kohler, 1947; Pomerantz & Kubovy, 1986). Objects often appear in different contexts and are almost never imaged from the same viewpoint; thus, the retinal images associated with physical objects are generally complex and varied. To have any hope of obtaining a useful interpretation of the retinal images, such as recognizing objects that have been encountered previously, there must be initial processes that organize the image data into those groups most likely to form meaningful objects.

Perceptual organization is also important because it generally results in highly compact representations of the images, facilitating later processing, storage, and retrieval. (See Witkin & Tenenbaum, 1983, for a discussion of the importance of perceptual organization from the viewpoint of computational vision.) Although much has been learned about the mechanisms of perceptual organization (see, e. g. , Beck, 1982; Bergen, 1991; Palmer & Rock, 1994; Pomerantz & Kubovy, 1986), progress in developing testable quantitative theories has been slow.

One area where substantial progress has been made is in models of texture grouping and segregation. These models have begun to put the study of perceptual organization on a firm theoretical footing that is consistent with the psychophysics and physiology of low-level vision. Two general types of model for texture segregation have been proposed. In the feature-based models, retinal images are initially processed by mechanisms that find specific features, such as edge segments, line segments, blobs, and terminators.

Grouping and segregation are then accomplished by finding the image regions that contain the same feature or cluster of features (see, e. g. , Julesz, 1984, 1986; Marr, 1982; Treisman, 1985). These models are relatively simple, are consistent with some aspects of low-level vision, and have been able to account for a range of experimental results. In the filter-based models, retinal images are initially processed by tuned channels, for example, " contrast-energy" channels selective for size and orientation.

Grouping and segregation are then accomplished by finding those image regions with approximately constant output from one or more channels (Beck, Sutter, & Ivry, 1987; Bergen & Landy, 1991; Bovik, Clark, & Geisler, 1990; Caelli, 1988; Chubb & Sperling, 1988; Clark, Bovik, & Geisler, 1987; Fogel & Sagi, 1989; Graham, Sutter, & Venkatesan, 1993; Victor, 1988; Victor & Conte, 1991; Wilson & Richards, 1992).

These models have some advantages over the existing feature-based models: They can be applied to arbitrary images, they are generally more consistent with known low-level mechanisms in the visual system, and they have proven capable of accounting for a wider range of experimental results.

However, the current models do not make accurate predictions for certain important classes of stimuli. One class of stimuli are those that contain regions of texture that can be segregated only on the basis of local structure (i. e. , shape).

Another broad class of stimuli for which most current perceptual organization models do not make adequate predictions are those containing nonstationary structures; specifically, structures that change smoothly and systematically across space. Nonstationary structures are the general rule in natural images because of perspective projection, and because many natural objects are the result of some irregular growth or erosion process. A simple example of a nonstationary structure would be a contour formed by a sequence of line segments (a dashed contour) embedded in a background of randomly oriented line segments.

Such contours are usually easily picked out by human observers. However, the elements of the contours cannot be grouped by the mechanisms contained in current filter-based or feature-based models, because no single orientation channel or feature is activated across the whole contour. Grouping the elements of such contours requires some kind of contour integration process that binds the successive contour elements together on the basis of local similarity. A more complex example of a nonstationary structure would be an image of wood grain.

Such a texture contains many contours whose spacing, orientation, and curvature vary smoothly across the image. Again, such textures are easily grouped by human observers but cannot be grouped by the mechanisms contained in the current models. Grouping the contour elements of such

textures requires some form of texture integration (the two-dimensional analogue of contour integration). The heart of the problem for existing quantitative models of grouping and segregation is that they do not represent the structure of the image data with the richness achieved by the human visual system.

The human visual system apparently represents image information in an elaborate hierarchical fashion that captures many of the spatial, temporal, and chromatic relationships among the entities grouped at each level of the hierarchy. Grouping and segregation based on simple feature distinctions or channel responses may well be an important initial component of perceptual organization, but the final organization that emerges must depend on more sophisticated processes.

The major theoretical aim of this study was to develop a framework for constructing and testing models of perceptual organization that capture some of the richness and complexity of the representations extracted by the human visual system, and yet are computationally well defined and biologically possible. Within this framework, we have developed a model of perceptual organization for two-dimensional (2D) line images and evaluated it on a number of " textbook" perceptual organization demonstrations.

In this article we refer to this model as the extended model when it is necessary to distinguish it from a simplified version, the restricted model, described later. Perceptual organization must depend in some way on detected similarities and differences between image elements. Furthermore, it is obvious that similarities and differences along many different stimulus dimensions can contribute to the organization that is perceived. Although

there have been many studies of individual stimulus dimensions, there have been few systematic attempts to study how multiple dimensions interact (Beck et al.

, 1987; Fahle & Abele, 1996; Li & Lennie, 1996). The major experimental aim of this study was to measure how multiple stimulus dimensions are combined to determine grouping strength between image elements. To this end, we conducted a series of three-pattern grouping experiments to directly measure the tradeoffs among two, three, or four stimulus dimensions at a time. Predictions for these experiments were generated by a restricted version of the model appropriate for the experimental task. The experimental results provided both a test for the restricted model and a means of estimating the model's parameters.

The estimated parameter values were used to generate the predictions of the extended model for complex patterns. The next four sections describe, respectively, the theoretical framework, the restricted model, the experiments and results, and the extended model and demonstrations. Theoretical Framework for Perceptual Organization In this section we discuss four important components of perceptual organization: hierarchical representation, detection of primitives, detection of similarities and differences among image parts, and mechanisms for grouping image parts.

These components taken together form the theoretical framework on which the restricted and extended quantitative models are based. Hierarchical Representation It is evident that the mechanisms of perceptual organization yield a rich hierarchical representation that describes the relationship of " parts" to " wholes" at a number of levels; that is, the wholes at one level

often become the parts at the next level. However, there is evidence that the process by which the hierarchical representation is constructed does not proceed strictly either from local to global or from global to local.

The global structure of a large letter composed of small letters can be discovered before the structure of the individual small letters is discovered (Navon, 1977), and there exist ambiguous figures, such as R. C. James's classic Dalmatian dog, that can be solved locally only after at least some of the global structure is discovered. On the other hand, the discovery of structure must sometimes proceed from local to global; for example, it would be hard to extract the symmetry of a complex object without first extracting some of the structure of its subobjects.

Any well-specified theory of perceptual organization must define what is meant by parts, wholes, and relationships between parts and wholes. Given the current state of knowledge, all definitions, including the ones we have adopted, must be tentative. Nonetheless, some basic definitions must be made in order to form working models. In our framework, the most primitive objects are defined on the basis of the current understanding of image encoding in the primary visual cortex of the primate visual system.

Higher order objects are defined to be collections of lower order objects (which may include primitive objects), together with information about the relationships between the lower order objects. The range of relationships that the visual system can discover, the order and speed with which they are discovered, and the mechanisms used to find them are unsettled issues. As a starting point the relationships we consider are quantitative similarities and differences in size, position, orientation, color, and shape.

These dimensions were picked for historical and intuitive reasons: They are major categories in human language and therefore are likely to correspond to perceptually important categories. The precise definitions of these dimensions of similarity between objects are given later. Detection of Primitives: Receptive-Field Matching One of the simplest mechanisms for detecting structure within an image is receptive-field matching, in which relatively hard-wired circuits are used to detect the different spatial patterns of interest.

For example, simple cells in the primary visual cortex of monkeys behave approximately like hard-wired templates: A strong response from a simple cell indicates the presence of a local image pattern with a position, orientation, size (spatial frequency), and phase (e. g. , even or odd symmetry) similar to that of the receptive-field profile (Hubel & Wiesel, 1968; for a review, see DeValois & DeValois, 1988). The complex cells in the primary visual cortex are another example.

A strong response from a typical complex cell indicates a particular position, orientation, and spatial frequency independent of the spatial phase (Hubel & Wiesel, 1968; DeValois & DeValois, 1988). Receptive field matching may occur in areas other than the primary visual cortex, and may involve detection of image structures other than local luminance or chromatic contours, for example, structures such as phase discontinuities (von der Heydt & Peterhans, 1989) and simple radially symmetric patterns (Gallant, Braun, & Van Essen, 1993).

An important aspect of receptive-field matching in the visual cortex is that the information at each spatial location is encoded by a large number of

neurons, each selective to a particular size or scale. The population as a whole ps a wide range of scales and hence provides a " multiresolution" or " multiscale" representation of the retinal images (see, e. g. , DeValois & DeValois, 1988). This multiresolution representation may play an important role in perceptual organization.

For example, grouping of low-resolution information may be used to constrain grouping of high-resolution information, and vice versa. The quantitative models described here assume that receptive-field matching provides the primitives for the subsequent perceptual organization mechanisms. However, to hold down the complexity of the models, the receptive-field matching stage is restricted to include only units similar to those of cortical simple cells with small receptive fields. These units proved sufficient for the line pattern stimuli used in the experiments and demonstrations.

Receptive-field matching is practical only for a few classes of simple image structure, such as contour segments; it is unreasonable to suppose that there are hard-wired receptive fields for every image structure that the visual system is able to detect, because of the combinatorial explosion in the number of receptive-field shapes that would be required. Thus, there must be additional, more flexible, mechanisms for detecting similarities and differences among image regions. These are discussed next. Similarity/Difference Detection Mechanisms

Structure exists within an image if and only if some systematic similarities and differences exist between regions in the image. Thus, at the heart of any perceptual organization system there must be mechanisms that match or

compare image regions to detect similarities and differences. (For this discussion, the reader may think of image regions as either parts of an image or as groups of detected primitives.) Transformational matching A well-known general method of comparing image regions is to find out how well the regions can be mapped onto each other, given certain allowable transformations (see, e.

g. , Neisser, 1967; Pitts & McCulloch, 1947; Rosenfeld & Kak, 1982; Shepard & Cooper, 1982; Ullman, 1996). The idea is, in effect, to use one image region as a transformable template for comparison with another image region. If the regions closely match, following application of one of the allowable transformations, then a certain similarity between the image regions has been detected. Furthermore, the specific transformation that produces the closest match provides information about the differences between the image regions.

For example, consider an image that contains two groups of small line segment primitives detected by receptive-field matching, such that each group of primitives forms a triangle. If some particular translation, rotation, and scaling of one of the groups brings it into perfect alignment with the other group then we would know that the two groups are identical in shape, and from the aligning transformation itself we would know how much the two groups differ in position, orientation, and size. There are many possible versions of transformational matching, and thus it represents a broad class of similarity-detection mechanisms.

Transformational matching is also very powerful—there is no relationship between two image regions that cannot be described given an appropriately

general set of allowable transformations. Thus, although there are other plausible mechanisms for detecting similarities and differences between image regions (see section on attribute matching), transformational matching is general enough to serve as a useful starting point for developing and evaluating quantitative models of perceptual organization. Use of both spatial position and color

The most obvious form of transformational matching is based on standard template matching; that is, maximizing the correlation between the two image regions under thefamilyof allowable transformations. However, template matching has a well-known limitation that often produces undesirable results. To understand the problem, note that each point in the two image regions is described by a position and a color. The most general form of matching would consist of comparing both the positions and colors of the points. However, standard template matching compares only the colors (e. g. , gray levels 2) at like positions.

If the points cannot be lined up in space then large match errors may occur even though the positional errors may be small. A more useful and plausible form of matching mechanism would treat spatial and color information more equivalently by comparing both the spatial positions and the colors of the points or parts making up the objects. For such mechanisms, if the colors of the objects are identical then similarity is determined solely by how well the spatial coordinates of the points or parts making up the objects can be aligned and on the values of the spatial transformations that bring them into the best possible alignment.

In other words, when the colors are the same, then the matching error is described by differences in spatial position. For such mechanisms, B matches A better than B matches C, in agreement with intuition. Later we describe a simple matching mechanism that simultaneously compares both the spatial positions and the colors of object points. We show that this mechanism produces matching results that are generally more perceptually sensible than those of template matching. Attribute matching

Another well-known method of comparing groups is to measure various attributes or properties of the groups, and then represent the differences in the groups by differences in the measured attributes (see, e. g. , Neisser, 1967; Rosenfeld & Kak, 1982; Selfridge, 1956; Sutherland, 1957). These attributes might be simple measures, such as the mean and variance of the color, position, orientation, or size of the primitives in a group, or they might be more complex measures, such as the invariant shape moments. It is likely that perceptual organization in the human visual system involves both transformational matching and attribute matching.

However, the specific models considered here involve transformational matching exclusively. The primary reason is that perceptual organization models based on transformational matching have relatively few free parameters, yet they are sensitive to differences in image structure—an essential requirement for moving beyond existing filter- and feature-based models. For example, a simple transformational matching mechanism (described later) can detect small differences in arbitrary 2D shapes without requiring an explicit description of the shapes.

On the other hand, specifying an attribute-matching model that can detect small differences in arbitrary shapes requires specifying a set of attributes that can describe all the relevant details of arbitrary shapes. This type of model would require many assumptions and/or free parameters. Our current view is that transformational matching (or something like it) may be the central mechanism for similarity/difference detection and that it is supplemented by certain forms of attribute matching. Matching groups to categories

The discussion so far has assumed implicitly that transformational and attribute matching occur between different groups extracted from the image. However, it is obvious that the brain is also able to compare groups with stored information because this is essential for memory. Thus, the visual system may also measure similarities and differences between groups and stored categories, and perform subsequent grouping using these similarities and differences. These stored categories might be represented by prototypes or sets of attributes.

Rather than use stored categories, the visual system could also measure similarities and differences to categories that emerge during the perceptual processing of the image. For example, the visual system could extract categories corresponding to prevalent colors within the image, and then perform subsequent grouping on the basis of similarities between the colors of image primitives and these emergent color categories. Grouping Mechanisms Once similarities and differences among image parts are discovered, then the parts may be grouped into wholes.

These wholes may then be grouped to form larger wholes, resegregated into a different collection of parts, or both. However, it is important to keep in mind that some grouping can occur before all of the relevant relationships between the parts have been discovered. For example, it is possible to group together all image regions that have a similar color, before discovering the geometrical relationships among the regions. As further relationships are discovered, the representations of wholes may be enriched, new wholes may be formed, or wholes may be broken into new parts and reformed.

Thus, the discovery of structure is likely to be an asynchronous process that operates simultaneously at multiple levels, often involving an elaborate interleaving of similarity/difference detection and grouping. Within the theoretical framework proposed here we consider one grouping constraint—the generalized uniqueness principle—and three grouping mechanisms: transitive grouping, nontransitive grouping, and multilevel grouping. The uniqueness principle and the grouping mechanisms can be applied at multiple levels and can be interleaved with similarity/difference detection.

Generalized uniqueness principle The uniqueness principle proposed here is more general: it enforces the constraint that at any time, and at any level in the hierarchy, a given object (part) can be assigned to only one superordinate object (whole). An object at the lowest level (a primitive) in the hierarchy can be assigned to only one object at the next level, which in turn can be assigned to only one object at the next level, and so on. The sequence of nested objects in the hierarchy containing a given object is called the part–whole path of the object.

The generalized uniqueness principle, if valid, constrains the possible perceptual organizations that can be found by the visual system. Nontransitive grouping Our working hypothesis is that similarity in spatial position (proximity) contributes weakly to nontransitive grouping. If proximity were making a dominant contribution, then separated objects could not bind together separately from the background objects. Proximity contributes powerfully to a different grouping mechanism, transitive grouping, which is described next.

We propose that transitive and nontransitive grouping are in some competition with each other and that the visual system uses both mechanisms in the search for image structure. References Beck, J. (Ed.). (1982). Organization and representation in perception. Hillsdale, NJ: Erlbaum. Beck, J. , Sutter, A. , & Ivry, R. (1987). Spatial frequency channels and perceptual grouping in texture segregation. Computer Vision, Graphics and Image Processing, 37, 299–325. Bergen, J. R. (1991). Theories of visual texture perception. In D. Regan (Ed.), Spatial vision (pp. 114–134). New York: Macmillan. Bergen, J. R.

, & Landy, M. S. (1991). Computational modeling of visual texture segregation. In M. S. Landy & J. A. Movshon (Eds.), Computational models of visual processing (pp. 253–271). Cambridge, MA: MIT Press. Bovik, A. C. , Clark, M. , & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, 55–73. Caelli, T. M. (1988). An adaptive computational model for texture segmentation. IEEE Transactions on Systems, Man and

Cybernetics, 18, 9–17. Chubb, C. , & Sperling, G. (1988). Processing stages in non-Fourier motion perception.

Investigative Ophthalmology and VisualScience, 29Suppl. 266. Clark, M. , Bovik, A. C. , & Geisler, W. S. (1987). Texture segmentation using a class of narrowband filters. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 571–574). New York: IEEE. Fahle, M. , & Abele, M. (1996). Sub-threshold summation of orientation, color, and luminance cues in figure–grounddiscrimination. Investigative Ophthalmology and Visual Science, 37Suppl. S1147. Fogel, I. , & Sagi, D. (1989). Gabor filters as texture discriminator. Biological Cybernetics, 61, 103–113.

Gallant, J. L. , Braun, J. , & Van Essen, D. C. (1993, January). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. Science, 259, 100–103. Geisler, W. S. , & Albrecht, D. G. (1995). Bayesian analysis of identification in monkey visual cortex: Nonlinear mechanisms and stimulus certainty. Vision Research, 35, 2723–2730. Geisler, W. S. , & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination and identification. Visual Neuroscience, 14, 897–919. Geisler, W. S. , & Chou, K. (1995). Separation of low-level and high-level fac