Further Readings

Halle, M., and S. J. Keyser. (1998). Robert Frost's loose and strict iambics. In E. Iwamoto, M. Muraki, M. Tokunaga, and N. Hasegawa, Eds., *Festschrift in Honor of Kazuko Inoue.* Tokyo, Japan: Kanda University of International Studies.

Mid-Level Vision

Mid-level vision refers to a putative level of visual processing, situated between the analysis of the image (lower-level vision) and the recognition of specific objects and events (HIGH-LEVEL VISION). It is largely a viewer-centered process, seemingly concerned explicitly with real-world scenes, not simply images (see Nakayama, He, and Shimojo 1995). Yet, in distinction to high-level vision, mid-level vision represents the world only in a most general way, dealing primarily with surfaces and objects and the fact that they can appear at different orientations, can be variously illuminated, and can be partially occluded.

Vision as we understand it today is far more complicated than had been recognized even thirty to forty years ago. Despite the seeming unity of our visual experience, there is mounting evidence that vision is not a single function but is likely to be a conglomerate of functions, each acting with considerable autonomy (Goodale 1995; Ungerleider and Mishkin 1982). Along with this new appreciation of vision's complexity comes the striking fact that from a purely anatomical point of view, the portion of the brain devoted to vision is also much greater than previously supposed (Allman and Kaas 1975; Felleman and van Essen 1991). For example, about 50 percent of the CEREBRAL CORTEX of primates is devoted exclusively to visual processing, and the estimated territory for humans is nearly comparable. So vision by itself looms very large even when stacked up against all other conceivable functions of the brain. As such, subdivisions in vision, particularly principled ones that delineate qualitatively different processes, are sorely needed, and Marr's (1982) seminal argument for three levels provides the broad base for what we outline here.

Let us consider what processes might constitute mid-level vision, and then contrast them with low-level and high-level vision. Good examples of mid-level visual processing can be seen in the work of Kanizsa (1979). Compare figure 1a where we see many isolated fragments with figure 1b where the same fragments are accompanied by additional diagonal line segments. In figure 1b there is a dramatic shift in what is perceived. The isolated pieces seen in figure 1a now form a single larger figure, the familiar Necker cube.

The phenomenon just described is characterized by several things, which all appear to be related to objects and surfaces and their boundaries. Furthermore, they are examples of occlusion, the partial covering of one surface by another. There is also the indication of inferences being made, enabling us to represent something that has been made invisible. We are thus aware of something continuing behind, which in turn enables us to see a single figure, not isolated fragments.
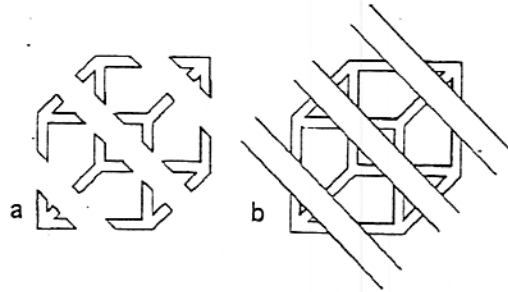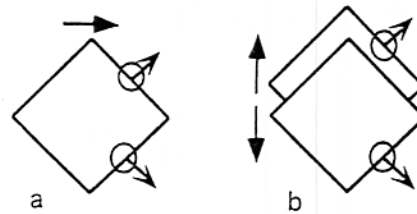

Figure 1.

Figure 2.

These characteristics, while not delineating mid-level vision in its entirety, provide sufficient basis for characterizing it as qualitatively different from low- and high-level vision. Consider the "aperture" problem for motion and its solution, something that until recently has been considered as within the province of low-level vision. Since Wallach's work (1935/1997), it has been recognized that there is an inherent ambiguity of perception if motion is analyzed locally, as would be the case for directionally selective receptive fields (see circles in figure 2). Thus in the case of a rightward-moving diamond (figure 2a), the local motions of the edges are very different from the motion of the whole figure. Yet, we are unaware of these local motions and see unified motion to the right. Computational models based on local motion measurements alone can recover the horizontal motion of the single figure on the left, but they cannot account for the perceived motion of one figure moving differently from another on the right (figure 2b). Although the local motions here are essentially identical, our visual system sees the motion in each case to be very different. It sees rightward motion of a single object versus opposing vertical motion of two objects. Only by the explicit parsing of the moving scene into separate surfaces can the correct motion be recovered. Thus, directionally selective neurons by themselves cannot supply reliable information regarding the motion of objects. Mid-level vision, with its explicit encoding of distinct surfaces, is required.

How might we distinguish mid-level from high-level vision? Consider figure 3. Most obvious is the reversal of the duck and the rabbit. From the above discussion, it should be clear that this reversal cannot be happening at the level of mid-level vision, which concerns itself more generally with surfaces and objects, but at higher levels where specific objects, like rabbits and ducks, are represented. For mid-level vision there is *no* reversal. Here mid-level vision's
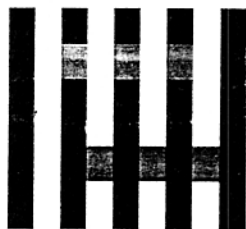
Figure 3.

Figure 4.

Figure 5.

job is to make sure we see a single thing or surface, despite its division into four separate image fragments by the overlying occluder and despite the change in its identity (the rabbit vs. the duck).

Another job of mid-level vision is to cope effectively with the characteristics of reflected light as it plays across surfaces in natural scenes. Surfaces can appear in various guises in the image, the result of being illuminated from various angles, being shaded by themselves or other surfaces, and by being viewed through transparent media. It would thus seem natural that various visual mechanisms would have developed or evolved to deal with these issues of illumination just as they have for cases of occlusion. This view is strengthened by the existence of perceptual phenomena that provide at least some hint as to how such processes may be occurring, also demonstrating the existence of processing that cannot be explained by low-level vision, say by lateral interactions of neurons with various types of receptive fields. Consider White's illusion shown in figure 4 where the apparent difference in brightness of the gray squares (top vs. bottom row) is very large despite being of equal luminance. Each identical gray patch is bounded by identical amounts of black and white areas, thus ruling out any explanation based on simultaneous contrast or lateral inhibition. The major difference is the nature of the junction structure bounding the areas, properties very important in mid-level vision processing. Figure 5 suggests that mid-level vision's role is the processing of shadows, showing how specific are the requirements for a dark region to be categorized as shadow and how consequential this categorization is for higher-level recognition. On the left we see a 3-D figure, a face. On the right, it looks more 2-D, where the outline around the dark region diminishes the impression that the figure contains shadows.

Although phenomena related to mid-level vision have been well-known, starting with GESTALT PSYCHOLOGY and more recently with work by Kanizsa (1979), the scope and positioning of mid-level vision in the larger scheme of visual processing has been unclear. Recently, Nakayama et al. (1995) have suggested that mid-level vision, in the form

of surface representation, is required for a range of processes more traditionally associated with early vision, including motion perception (see MOTION, PERCEPTION OF), forms of stereopsis, TEXTURE segregation and saliency coding. More speculatively, there has been a proposal that mid-level vision is the first level of processing, the results of which are available to conscious awareness (Jackendoff 1987; Nakayama, He, and Shimojo 1995), thus implying that mid-level vision is the earliest level to which ATTENTION can be deployed.

See also CONSCIOUSNESS; GESTALT PERCEPTION; ILLUSIONS; SHAPE PERCEPTION; SURFACE PERCEPTION; VISUAL PROCESSING STREAMS

—Ken Nakayama

References

Allman, J. M., and J. H. Kaas. (1975). The dorsomedial cortical visual area: a third tier area in the occipital lobe of the owl monkey (*Aotus trivirgatus*). *Brain Research* 100: 473–487.

Felleman, D. J., and D. C. van Essen. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1: 1–47.

Goodale, M. A. (1995). The cortical organization of visual perception and visuomotor control. In S. M. Kosslyn and D. N. Osherson, Eds., *Visual Cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.

Kanizsa, G. (1979). *Organization in Vision: Essays on Gestalt Perception*. New York: Praeger.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Nakayama, K., Z. J. He, and S. Shimojo. (1995). Visual surface representation: a critical link between lower-level and higher-level vision. In S. M. Kosslyn and D. N. Osherson, Eds., *Visual Cognition*. Cambridge, MA: MIT Press, pp. 1–70.

Ungerleider, L. G., and M. Mishkin. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, Eds., *Analysis of Visual Behavior*. Cambridge, MA: MIT Press.

Wallach, H. (1935). Über visuell wahrgenommene Bewegungsrichtung. *Psychol. forschung* 20: 325–380. Translated (1997) by S. Wuenger, R. Shapley, and N. Rubin. On the visually perceived direction of motion. *Perception* 25: 1317–1368.

Mind-Body Problem

The *mind-body problem* is the problem of explaining how our mental states, events, and processes are related to the physical states, events, and processes in our bodies. A ques-