Optic flow and depth perception

WILLIAM A. SIMPSON

Department of Psychology, University of Winnipeg, 515 Portage Avenue, Winnipeg, Manitoba, Canada R3B 2E9

Received 16 December 1991; revised 18 May 1992; accepted 20 May 1992

Abstract—The field of depth recovery from optic flow has recently experienced much growth, both on the theoretical and on the empirical fronts. Unfortunately, the theoretical results are not as widely known to perception workers as they might be. This article gives a simple analysis of the information for depth present in optic flow. It also reviews the psychophysical results for depth recovery from motion. These results are discussed with reference to the theoretical analysis and to relevant computer algorithms for depth recovery.

In his tribute to Hermann von Helmholtz, James Clerk Maxwell (1877) nicely stated the plight of the perception worker:

In no department of research is the combined and concentrated light of all the sciences more necessary than in the investigation of sensation (p. 390).

Maxwell’s observation is especially true for the study of depth perception through motion. As in all other areas of perception, knowledge of the relevant physics and physiology is required in addition to the perceptual literature. In depth from motion, however, the physics is rather complicated (3-dimensional [3D] kinematics and vector analysis), our knowledge of the physiology of visual motion detection is advancing rapidly, there is a large and technically demanding computer vision literature, and the literatures for the psychophysics of motion, depth (stereopsis) and depth from motion must all be mastered. An exhaustive review of all this literature is perhaps beyond even the capabilities of a Helmholtz. I will, however, attempt to give an adequate background for the field of depth perception through motion. I will first briefly describe the phenomenon of depth perception from motion. Next I will analyze the depth information available from visual motion. The balance of the paper is concerned with various issues in the recovery of depth from motion by humans, with references to relevant computer vision work.

1. THE PHENOMENON OF DEPTH FROM MOTION

Before we consider the information present that makes it possible to get depth from motion, let us first describe the phenomenon. It is simply this: certain static scenes look flat; movement lets the viewer assign depths to the scene elements. Helmholtz (1925) gives an early statement of the effectiveness of motion as a source of depth information:

Suppose, for instance that a person is standing still in a thick woods, where it is impossible for him to distinguish, except vaguely and roughly, in the mass of foliage and branches all around him what belongs to one tree and what to another, or how far apart the separate trees are, etc. But the moment he begins to move forward,
everything disentangles itself, and immediately he gets an apperception of the material contents of the woods and their relations to each other in space, just as if he were looking at a good stereoscopic view of it (pp. 295–296).

Bourdon (1898) was one of the first to explore the perception of depth under conditions of relative motion between stimulus and observer. Since Bourdon there have been many laboratory demonstrations that a stimulus array seen as flat when stationary jumps out in depth when it is moved. Perhaps the most familiar of these demonstrations are those of Wallach and O'Connell (1953). Objects are laid out in depth on the surface of a turntable. The shadow of the rotating objects is projected onto a screen where a viewer watches the display. This method eliminates information from convergence, accommodation, and binocular disparity, yet viewers can readily reconstruct the layout of the objects.

The depth-from-motion phenomenon is analogous to Julesz's demonstrations of depth from binocular disparity. In the absence of disparity, no depth can be seen in Julesz's (1960, 1971) random-dot stereograms. In the absence of motion, no depth can be seen in Wallach and O'Connell's displays. Just as disparity is sufficient to yield depth, so is motion.

Many studies of this sort have been done, and they will be reviewed in the psychophysics section of this paper. Although Wallach and O'Connell generated their displays by moving real objects, many authors manipulate points on a cathode ray tube in any way that might generate an impression of depth. Manipulations of velocities on the face of the screen do not necessarily simulate the rigid motion of objects in 3D space. The analysis in Section 2 will be restricted to optic flow generated by rigid motion—that is, either all the environmental points are stationary and the observer is moving, or (equivalently) the observer is stationary and the points all move together as one.

Before we look at any more results from experiments on human perception of depth from motion, let us consider the information in the projection of moving points upon which this ability is based.

2. THE DEPTH INFORMATION IN OPTIC FLOW

As an observer moves relative to his environment, the rays connecting object points to the observer's vantage point move in a regular fashion. Gibson (Gibson, 1950; Gibson et al., 1955) was among the first to note the richness of information to be found in this pattern of angular velocities, which he called 'motion perspective' or simply 'flow'. The first instance of the term 'optical flow' that I can find is in Gibson (1966, p. 161). (I prefer 'optic flow' to 'optical flow' since the former is consistent with Gibson's 'optic array'.) When the optic flow is projected on the retina, the result is a field of retinal velocities called the retinal flow. The terminology in this field is not altogether standardized; some authors use 'optic flow' to refer to what I have just defined as retinal flow. However the distinction is useful.

Many sophisticated analyses of optic (or retinal) flow have been done. Important papers include those of Longuet-Higgins and Prazdny (1980), Prazdny (1983), Koenderink (1986), and Waxman and Wohn (1988); for an analysis along quite different lines, see Lappin (1990). The basic assumption of all analyses is that the observer moves relative to a rigid environment. That is, either the observer moves amongst static objects or (identically) all the objects move together as one relative to the