Sensor fusion in motion perception[1]

David Coombs
david.coombs@nist.gov; Robot Systems Division, National Institute of Standards and Technology, Gaithersburg, MD 20899

A goal of robotics is to discover principles that enable systems to behave robustly in complex domains, and animals offer rich examples of such systems to study. Wertheim's theory elegantly unifies empirical data and existing theories. It is satisfying to see the theory grounded in the neurophysiological structures that have so far been implicated, since implementation feasibility is important. Nevertheless, speculation on additional cues to motion perception might be reasonable. Vision is the primary cue to perceiving motion of objects, but there are several possible clues to ego motion (motor efference from eye, neck, and body, vestibular and otolith afference, oculomotor kinesthesis, visual afference, and expectations) and these may influence the perception of motion generally if they help inform ego-motion estimation. Ideally, a robot would perceive motion and behave appropriately. The system would attempt to assimilate sensory data from several modalities to estimate the motion of the creature itself, and of nearby objects. Human motion perception can be considered from this viewpoint.

Suppose that one drives a car over a bumpy road. The largest visual flows indicate forward translation, and yet the vestibular and otolith systems do not. They indicate accelerations of the head with respect to a coordinate system (CS) that is a low-pass filtered (smoothed) version of the (bouncing) car CS, since the subject has reached a steady velocity and is no longer accelerating. There are three possibilities. The vestibular and otolith signals fully describe the subject's motion and a large segment of the world is moving rapidly past the subject; or there is a steady state component of the subject's motion to which the vestibular and otolith organs no longer respond; or both the subject and the scene are moving with respect to the world. How does the system choose a combination of ego motion and object motion that accounts for the observations?

Two obvious policies are to take the most likely interpretation based on experience, and to make the most conservative judgment with respect to the creature's safety. The creature must presume that it can influence its motion, so it might as well attribute perceived coherent large-field motion to ego motion and behave accordingly. If it cannot control the relative motion, then its response will be irrelevant, but if it can, it might avoid colliding with trees, for instance. It might be argued, then, that the safe interpretation is to attribute the recent history of sensory cues to ego motion.

A crucial element of Wertheim's model is the incorporation of the sensory data histories and the resulting interactions. The temporal characteristics of each cue might be considered. What determines the spatiotemporal properties of the optokinetic signal that contributes to perception of ego motion? Certainly the range of image flows that can be used is limited by the range that can be perceived, but is use of the data further limited? It has been noted [Howard, 1982] that the sense of vection approximately follows the time course of the decay of the vestibular system's response to constant rotation of the body. These spatiotemporal response characteristics of complementary signals such as vestibular data might be natural for limiting the use of visual signals to charge up the eye ego-motion reference signal. Suppose that image flows corresponding to ego motions below the sensitivity of the vestibular system contribute to the reference signal immediately and faster flows contribute to the reference signal only after the vestibular system could be expected to fall silent. Then vection would be induced immediately at accelerations below the sensitivity of the vestibular system. Vection would result later for larger accelerations only after the history of conflicting vestibular data is sufficiently old that the subject may have been slowly accelerated, undetected by the vestibular system. Thus, motion would be attributed to both the subject and the scene until enough time has passed that the subject's motion alone could account for the visual motion.

Consider the extinction of vection in an ordinary situation. Suppose that one is sitting in a train and an adjacent train begins to move, inducing vection. When one looks up and sees that another train (and not one's own) is moving, the vection is extinguished. What factors could have extinguished vection? It could not have been vestibular sensations resulting from looking up, or they would also suppress the veridical perception of ego motion when one's train does begin to move. Again, there are three cases. The other train moves, one's own train moves, and both trains coincidentally begin to move at the same time. When one looks out the window, one can presumably determine whether either or both of the trains are moving against the ground and trees. (It might be interesting to know whether there is a difference in the reaction times to extinguish vection and to confirm veridical ego motion.) This example suggests that some sensory fusion occurs at a high level even if it is not necessarily required for behavioral responses. It further confirms that in ordinary behavior, the creature may actively seek out additional information to assess the situation. There is a related question I would like to consider. When a person experiences vection in a parked car, a common reaction seems to be stomping on the brake pedal before looking up to determine whether the car is in fact rolling. Does a train passenger ever stab at an imaginary brake pedal at the onset of vection? If not, then this suggests that low level "reflexive" reactions to ego-motion perception are influenced by high level contextual expectations. It may be impossible to decouple models of motion perception from models of behavior and expectation; it seems likely that higher-level models will be necessary to predict many observed behaviors.

References

[Howard, 1982] Howard, I. (1982). *Human Visual Orientation.* Wiley & Sons.

[1]This commentary is contained in the target article "Motion perception during self-motion: The direct versus inferential controversy revisited," by A. H. Wertheim, *BEHAVIORAL AND BRAIN SCIENCES* (1994) 17:2, pp. 293–355