Vision Research 107 (2015) 101–112

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The effective reference frame in perceptual judgments of motion direction

Mehmet N. Agaoglu [a,b,*], Michael H. Herzog [c], Haluk Öğmen [a,b]

[a] Department of Electrical and Computer Engineering, University of Houston, N308 Engineering Building 1, Houston, TX 77204-4005, USA
[b] Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX 77204-4005, USA
[c] Laboratory of Psychophysics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 13 May 2014
Received in revised form 26 November 2014
Available online 20 December 2014

Keywords:
Motion perception
Reference frames
Eye movements

ABSTRACT

The retinotopic projection of stimulus motion depends both on the motion of the stimulus and the movements of the observer. In this study, we aimed to quantify the contributions of endogenous (retinotopic) and exogenous (spatiotopic and motion-based) reference frames on judgments of motion direction. We used a variant of the induced motion paradigm and we created different experimental conditions in which the predictions of each reference frame were different. Finally, assuming additive contributions from different reference frames, we used a linear model to account for the data. Our results suggest that the effective reference frame for motion perception emerges from an amalgamation of motion-based, retinotopic and spatiotopic reference frames. In determining the percept, the influence of relative motion, defined by a motion-based reference frame, dominates those of retinotopic and spatiotopic motions within a finite region. We interpret these findings within the context of the Reference Frame Metric Field (RFMF) theory, which states that local motion vectors might have perceptual reference-frame fields associated with them, and interactions between these fields determine the selection of the effective reference frame.

1. Introduction

When an object moves during steady fixation, its projection on the retina also moves at a speed proportional to its physical speed. The perceptual system readily interprets this retinal motion as the motion of an object in the environment. However, when the observer's eyes, head or body move, the retinal image motion does not directly correspond to a corresponding motion in the environment. In order to perceive veridically the motion of an object in the environment, the perceptual system needs to carry out coordinate transformations (Swanston, Wade, & Day, 1987; Wade & Swanston, 1987). In other words, the retinal motion due to self-motion or movement of the eyes need to be parsed out such that what is left directly corresponds to the motion of an object in the environment. Gibson argued that optic flow alone is sufficient to make the required transformations and to decompose retinal motion into self-motion and object motion relative to the scene (Gibson, 1979). Many psychophysical (e.g. Rushton, Bradshaw, & Warren, 2007; Warren & Rushton, 2009), neurophysiological (Duffy & Wurtz, 1991a; Duffy & Wurtz, 1991b), functional imaging (e.g. Morrone et al., 2000), and modeling (Furman & Gur, 2003; Pack, Grossberg, & Mingolla, 2001) studies supported his position. However, early studies of motion perception during smooth-pursuit eye movements showed that the coordinate transform from retinocentric reference frame to head-centric one is not perfect. A stationary object is perceived to be moving in the direction opposite to the direction of the ongoing pursuit eye-movement (Filehne illusion. Filehne, 1922; Freeman & Banks, 1998; Mack & Herman, 1972; Mack & Herman, 1973; Wertheim, 1987) and a moving object is perceived to be slower when it is tracked than when it is viewed during fixation (Aubert-Fleischl effect. Fleischl, 1882; Aubert, 1886; Freeman & Banks, 1998). The perceived direction and the extent of motion of an object that moves non-collinearly with the pursuit target significantly deviate from corresponding physical quantities (Becklen, Wallach, & Nitzberg, 1984; Festinger, Sedgwick, & Holtzman, 1976; Furman & Gur, 2005; Kano & Hayashi, 1981; Souman, Hooge, & Wertheim, 2005; Souman, Hooge, & Wertheim, 2006b). Assuming perfect retinal gains (i.e. the ratio of perceived and actual retinal motion extents or speeds is 1), these perceptual errors and illusions have been conventionally attributed to an under-registration of eye velocities. However, perceived retinal motion is strongly modulated by stimulus properties such as spatial frequency, dot density, contrast,

* Corresponding author at: Department of Electrical and Computer Engineering, University of Houston, N308 Engineering Building 1, Houston, TX 77204-4005, USA.

http://dx.doi.org/10.1016/j.visres.2014.12.009
0042-6989/

stimulus scale and chromatic content (see review by Nishida (2011)) and hence, errors in estimating retinal motion should also be considered in the computations of head-centric motions (Freeman & Banks, 1998).

Many models of motion perception during smooth pursuit have been proposed to quantify the degree to which this coordinate transformation is complete. In most of these models, the observer's head and body are assumed to be stationary with respect to the outside world, and the perceived head-centered motion is a combination of retinal motion and eye velocity estimates (Freeman, 2001; Freeman & Banks, 1998; Souman, Hooge, & Wertheim, 2006a; Swanston et al., 1987; Turano & Massof, 2001; Wertheim, 1994). Models with non-linear motion transducers have been shown to perform slightly better than those with linear motion estimators for both terms (Freeman, 2001; Turano & Massof, 2001). The estimated eye velocity in some of these models is a function of both retinal and extra-retinal signals, whereas retinal motion estimates depend only on stimulus parameters and retinal motion itself (Freeman & Banks, 1998; Turano & Massof, 2001; Wertheim, 1994). Several studies have concluded that perceived motion during pursuit also depends on stimulus parameters including size (Turano & Heidenreich, 1999), spatial frequency (Freeman & Banks, 1998; Wertheim, 1994), speed (Pola & Wyatt, 1989; Turano & Heidenreich, 1996), and presentation duration (Mack & Herman, 1978; Souman et al., 2005; Wertheim, 1987).

When there are two objects in the scene and one of them is tracked, the relative motion between the objects may become a major determinant of perceived motion. In some studies, this fact was overlooked and the failure to discriminate relative motion from retinal motion led some researchers to conclude that the perceptual system has very weak (i.e. gains < 0.1) or no information at all about the ongoing pursuit eye movement (Dodge, 1904; Festinger et al., 1976; Stoper, 1973). For instance, when a small dot is pursued in a dark room and the motion of another (moving or stationary) dot is judged, the retinal motion of the target dot and its relative motion with respect to the pursuit dot are almost identical (assuming perfect smooth pursuit). It is impossible to decouple contributions of the retinal and relative motions in these displays. In fact, Mack and Herman (1978) showed that the relative motion between the pursuit target and the background object is one of the main factors influencing perceived motion. The contribution of the relative motion between the pursuit target and the background has been noted in several studies (Baker & Braddick, 1982; Brenner & van den Berg, 1994; Freeman, Champion, Sumnall, & Snowden, 2009; Freeman, Champion, & Warren, 2010; Hisakata, Terao, & Murakami, 2013; Mack & Herman, 1978; Mateeff, Hohnsbein, & Ehrenstein, 1990; Snowden, 1992; Turano & Heidenreich, 1999; Wallach, 1959; Wallach, O'Leary, & McMahon, 1982). In these studies, qualitative descriptions of how and when relative motion between the pursuit target and the background affects perceived motion have been given. Baker and Braddick (1982) argued that, at slow speeds, relative motion determines percepts whereas at high speeds, absolute motion (i.e. motion with respect to a spatiotopic reference frame such as stimulus display) takes over. Mack and Herman (1978) concluded that object-relative motion is only effective when the object of interest is in close proximity of the pursuit target. Brenner and van den Berg (1994) reported that the perceived target velocity does not change as long as the relative motion of the pursuit target with respect to a textured background is kept fixed.

When there are multiple moving objects in the scene, a typical scenario in normal viewing conditions, relative motions of these objects can fully determine the perceived motion. Duncker (1929) used displays generated by point-lights attached to an otherwise invisible rotating and translating circular cardboard (Duncker, 1929, pg. 240). When a point-light is attached to the rim of the cardboard, observers perceive cycloidal motion of the light, which corresponds to its trajectory on the retina if the observer's eyes are stationary. Percepts do not change when the point-light is tracked. However, when another point-light is added to the hub of the wheel, the central light is perceived to be translating linearly, whereas the peripheral light is perceived as rotating around the central light, regardless of whether the central light is tracked or not. In the latter case, the retinal trajectory of the point-light at the rim is again a cycloid; but the percepts dominantly correspond to its relative motion with respect to the central light. Similar and more complex demonstrations of the superiority of relative motion were done Johansson (1950) and Johansson (1973). In line with this, it has been shown that the thresholds for detecting relative motion is much less than those for absolute motion (Snowden, 1992). Moreover, the movements of the eyes, head or body result in relative motions of objects at different depths in the environment. A complete theory of motion perception, therefore, must take into account the relative motion of objects with respect to each other. Wade and Swanston's quantitative model of motion perception (Wade & Swanston, 1987) explicitly includes a term for relative motion of objects with respect to each other. According to their model, the registered retinal motion undergoes a sequence of coordinate transforms to reach a geocentric representation. Estimated retinal motions are compensated for estimated eye movements at the orbital level, and the output of this process is combined with the "pattern-centric" signals (i.e. relative motion). Furthermore, they proposed that the two signals are not treated equally, but each has a weight. A similar approach was taken by Gogel (1977). He also argued that the relative motion has a greater weight compared to the other components (Gogel, 1977). Unfortunately, the weights of different terms have never been determined experimentally.

In contrast to the models of motion perception mentioned so far, we adopted a top-down approach and modeled the perceived motion as an interplay between various reference frames available to the perceptual system. By doing so, we remained agnostic as to how coordinate transforms outlined by previous models take place; instead, we sought to investigate how the perceptual system forms the "effective reference frame". Let's assume that the head is kept still and two objects are moving in the fronto-parallel plane at different velocities. The perceived motion of each object depends on its motion on the retina (i.e. retinocentric or retinotopic reference frame), its motion on the display (i.e. space-centric or spatiotopic reference frame), and its motion relative to the motion of the other object (i.e. object-based or motion-based reference frame). The proposed model is given by

$$\text{Perceived motion} = w_s(d,\varphi)P_s + w_r(d,\varphi)P_r + w_{mb}(d,\varphi)P_{mb} + c, \tag{1}$$

where P_s, P_r, and P_{mb} represent the motion signals on spatiotopic, retinotopic and motion-based reference frames, and w_s, w_r, and w_{mb} represent the weights of each reference frame, respectively. The constant term c in the model captures the response bias of observers. The response bias represents byproducts of decision processes. Each P value represents also the predicted perceived-motion from a given reference frame. For instance, if observers perceive the motion direction solely based on retinal motion, (i.e. $w_s = 0$, $w_{mb} = 0$, and $w_r = 1$), perceived motion would be equal to P_r. Note that each weight is modeled as a function of distance d between the two objects and some other potential factors φ (such as perceptual groupings, stimulus scale, attention, etc.).

Equation (1) contains four unknowns, namely the three weights and the constant term. In order to have a unique solution, at least four linearly independent equations (i.e. different combinations of P_s, P_r, and P_{mb} values) are needed. To this end, we designed four