Perception, 1992, volume 21, pages 69–76

Gravity as a monocular cue for perception of absolute distance and/or absolute size

John S Watson ¶, Martin S Banks ¶ #, Claes von Hofsten †, Constance S Royden #
¶ Department of Psychology, University of California at Berkeley, Berkeley, CA 94720, USA
# School of Optometry, University of California at Berkeley, Berkeley, CA 94720, USA
† Department of Psychology, Umeå Universitet, Umeå, Sweden
Received 1 August 1990, in revised form 22 May 1991

Abstract. When the motion of an object is influenced by gravity (eg free fall, pendulum, wave motion), that influence may provide a cue to computing the absolute distance and/or size of the object. Formal analysis supports the claim that the distance and size of moving objects are generally computable with reference to the gravitational component of motion. Informal evidence from judgments of realism in films is consistent with this gravity-cue hypothesis.

1 Introduction

Most visual cues about the three-dimensional layout of the environment can only specify relative depth and relative size. Even the two most precise depth cues—binocular disparity and motion parallax (Lee 1974; Rogers and Graham 1982)—only provide distance information to within a scale factor. We need to judge absolute distance and size, however, in order to act in space. Take, for example, the act of catching a ball. For a perceiver to place his or her hands in the appropriate spatial position to intercept a ball and in an appropriate conformation to grasp it, he or she needs to know the absolute distance and size of the ball. If the ball is coming straight toward the eye, the expansion pattern of the ball on the retina will give the time-to-contact (Lee 1974). But in the more general case in which the ball is not heading toward the eye, the expansion pattern does not specify how far from the eye the ball is going to pass: it could be a distant large ball moving quickly or a near small ball moving slowly. Yet, the act of catching a ball is performed reliably by most people even under monocular viewing conditions (von Hofsten 1990).

Two well-known depth cues can specify absolute distance under the right circumstances: convergence and familiar size. Berkeley (1709) was the first to show that if interocular distance is known and the angle between the lines of sight of the two eyes is known, then the absolute distance to the point of convergence can in principle be derived. He argued that convergence was the basis for absolute distance perception. The usefulness of convergence as an absolute depth cue has been debated (eg Woodworth and Schlosberg 1954), but von Hofsten (1990) has shown that convergence can provide a precise scale for depth in near space. However, convergence is not an effective cue for objects at any appreciable distance and cannot account for performance under monocular viewing (von Hofsten 1990; Lie 1965).

In a monocular situation, the depth cue of familiar size could conceivably be used to calibrate space because if the true size and the retinal image size of an object are known, one can in principle estimate the absolute distance to the object.The depth perception of adults (Ittelson 1951) and infants (Yonas et al 1982) is influenced by familiar size. There is, however, a major problem with reliance on familiar size as the sole basis for absolute distance judgments. Many familiar shapes, such as spheres, or even people, come in a variety of sizes. Consequently, familiar size is unlikely to be a reliable source of information for guiding precise motor activity.

2 The gravity-cue hypothesis

We hypothesize that the motions of objects that are affected by gravity (eg projectile flight, pendulum motion, fluid wave motion) will supply important information about absolute distance and absolute size. In support of this hypothesis, we provide a quantitative argument that states that knowing the gravitational constant, which directly affects the trajectory and speed of projectile motion, allows one, in principle, to estimate the absolute distance to any object in free flight. Saxberg (1987a) has also evaluated this hypothesis and shown quantitatively that there is enough information in the trajectory and speed of projectile motion to estimate absolute distance. Later we consider the differences between his quantitative analysis and ours.

Consider the simplest case. A sphere is suspended in space. The retinal image in an observer's eye occupies 5 deg. The classic problem of estimating size and distance in this static scene results from the fact that there are infinite combinations of objective size and distance that can project a 5 deg image on the retina.

Consider a slightly more complex scene. The object moves laterally at some speed. If one does not known the objective speed, then the size and distance (and speed) remain ambiguous because there are infinite combinations of objective distance and speed that can produce a given rate of image movement on the retina.

Now imagine that the object's suspension is removed. If the observer perceives the motion and assumes it is governed by gravity, ambiguity can be resolved. In the simplest case of vertical fall, the first 0.25 s results in 1 ft of objective movement (ie descent $= 16t^2$). That fact can specify the size of the object. If the object falls a fifth of its diameter in the first 0.25 s of descent, it must have a diameter of 5 ft. Optical geometry would then allow the observer to resolve the distance to the object as being 57 ft [ie distance = descent/tan (deg subtended by descent), 57 = 1/tan(1 deg)].

Consider now the more complex and more common case of projectile motion involving components of motion laterally, vertically, and in depth. We will begin by showing that an analysis of the optic flow field, without regard to the action of gravity, cannot provide absolute distance information. Consider a stationary observer viewing an object moving on a linear path in an otherwise rigid scene. As shown in figure 1, we use a rectilinear coordinate system centered at the observer's eye. The Y axis is parallel to the gravitational vector. The object has the coordinates (X, Y, Z). The retina is represented by a two-dimensional plane, one unit behind the origin. The projection of the object has the retinal image coordinates (x, y), where $x = X/Z$ and $y = Y/Z$. The motion of the object is represented by three orthogonal vectors: V_x, V_y, and V_z for motions parallel to the X, Y, and Z axes, respectively.

From analyses of the optic flow field (Crowell et al 1989; Longuet-Higgins and Prazdny 1980):

$$x = -\frac{X_0 + V_x t}{Z_0 + V_z t}, \tag{1}$$

$$y = -\frac{Y_0 + V_y t}{Z_0 + V_z t}, \tag{2}$$

$$\boldsymbol{v}_x = \frac{V_z(X_0 + V_x t) - V_x(Z_0 + V_z t)}{(Z_0 + V_z t)^2}, \tag{3}$$

$$\boldsymbol{v}_y = \frac{V_z(Y_0 + V_y t) - V_y(Z_0 + V_z t)}{(Z_0 + V_z t)^2}, \tag{4}$$

$$\boldsymbol{a}_x = \frac{-2V_z V_x(Z_0 + V_z t) - 2V_z^2(X_0 + V_x t)}{(Z_0 + V_z t)^3}, \tag{5}$$

and

$$a_y = \frac{-2V_z V_y(Z_0 + V_z t) - 2V_z^2(Y_0 + V_y t)}{(Z_0 + V_z t)^3}, \tag{6}$$

$$\frac{Z}{V_z} = -\frac{x - x_e}{v_x}, \tag{7}$$

$$= -\frac{y - y_e}{v_y}, \tag{8}$$

where t is time, (x_e, y_e) are the retinal image coordinates of the so-called focus of expansion (the point from which the flow vectors streaming off the object seem to diverge or converge; see Longuet-Higgins and Prazdny 1980); v_x and v_y are the retinal image velocities, and a_x and a_y are the retinal accelerations in the directions x and y, respectively; V_x, V_y, and V_z are the real world velocities; and X_0, Y_0, and Z_0 are the coordinates of the initial position of the object. Equations (7) and (8) show that one can estimate the relative distance (Z/V_z) from retinal image properties. Without knowing the approach velocity, V_z, however, the absolute distance Z cannot be estimated from these equations. Examination of the other equations reveals that there are generally six unknown object properties (X, Y, Z, V_x, V_y, and V_z). With six equations [eg equations (1)-(4), (7), and (8)], one might expect to be able to solve for the six unknowns and thereby estimate the absolute distance of the object. The equations are, however, not independent and do not allow a solution for Z. Thus, as pointed out by Gibson et al (1955), Lee (1974), and Longuet-Higgins and Prazdny (1980), the optic flow field can provide estimates of relative but not absolute distance to objects moving at constant velocity.

The motion of objects, however, is frequently influenced by the action of gravity (as in projectile, hinged, or wave motion). In the case of projectile motion, the absolute

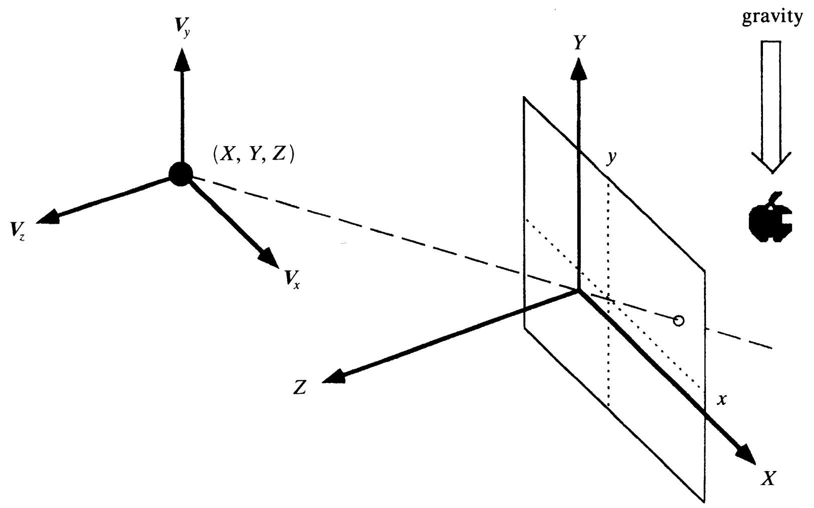

Figure 1. Coordinate system for a stationary observer and a moving object. The observer is represented by the three-dimensional coordinate system XYZ. The direction Y is parallel to gravity. The origin of this coordinate system is the center of the optics of the eye. The retina is represented by a plane, 1 unit behind the origin. The object has the coordinates (X, Y, Z) and moves in three directions with velocities represented by the vectors V_x, V_y, and V_z. The projection of the object onto the retina has the coordinates (x, y). Throughout this paper we use the convention that variables external to the observer are uppercase letters and variables internal to the observer (that is, retinal variables) are lowercase letters.

distance of an object can be found if the observer knows the gravitational constant and the direction of the gravitational vector. We disregard the effects of air friction (a notable exclusion, yet of diminishing importance for heavier objects and any object near apex), so the motion of the object is governed by its initial velocities, V_x, V_y, and V_z, and gravity only. Inclusion of the effects of gravity affects equations (1), (4), and (6) which become:

$$y = \frac{-Y_0 - V_y t + \frac{1}{2} g t^2}{(Z_0 + V_z t)} , \tag{9}$$

$$\boldsymbol{v}_y = \frac{V_z(Y_0 + V_y t - \frac{1}{2} g t^2) - (V_y - gt)(Z_0 + V_z t)}{(Z_0 + V_z t)^2} , \tag{10}$$

$$\boldsymbol{a}_y = \frac{2V_z(V_y - gt)(Z_0 + V_z t) + g(Z_0 + V_z t)^2 - 2V_z^2(Y_0 + V_y t - \frac{1}{2} g t^2)}{(Z_0 + V_z t)^3} , \tag{11}$$

where g is the acceleration due to gravity. There are different ways to solve for Z. One solution involves equations (8)–(11), the equations for the Y component of motion. Using equations (8)–(10) to substitute for V_z, V_y, and y, and collecting terms, the following relationship is derived:

$$Z = \frac{g(y - y_e)}{\boldsymbol{a}_y(y - y_e) - 2\boldsymbol{v}_y^2} . \tag{12}$$

In principle then, an observer can deduce the absolute distance to an object by measuring four retinal image properties in the direction corresponding to the gravitational vector: acceleration, velocity, position of the projected object, and the vertical position of the focus of expansion.

As mentioned earlier, Saxberg (1987a) has presented a similar quantitative analysis. Illustrating it here with our symbols and coordinates, he showed that one can compute the distance Z from the second derivatives of x and y [equations (5) and (6) above]. He obtains the following:

$$Z_0 = \frac{g(2\boldsymbol{v}_x + \boldsymbol{a}_x t)}{2(\boldsymbol{a}_y \boldsymbol{v}_x - \boldsymbol{a}_x \boldsymbol{v}_y)} . \tag{13}$$

Saxberg's equation is different from ours in two respects. First, it contains terms for both vertical and horizontal motion in the retinal image plane, whereas our equation (12) contains terms for vertical motion only. Second, Saxberg's equation contains velocity, acceleration, time, and the gravitational constant, whereas ours also contains a term for the vertical position of the focus of expansion.

Chapman (1968) and Todd (1981) have also examined the information contained in the trajectory of a freely falling object. They have shown that the image velocity, $\boldsymbol{v}_y$, is constant (that is, $\boldsymbol{a}_y = 0$) for an object on a trajectory that will ultimately hit the observer's eye. When $\boldsymbol{v}_y$ increases over time ($\boldsymbol{a}_y > 0$), the object will pass over the observer's head, and when $\boldsymbol{v}_y$ decreases over time ($\boldsymbol{a}_y < 0$), the object will fall short. As in Saxberg's analysis, the observer has to estimate image acceleration in order to use this information.

A human observer's ability to judge acceleration is rather limited (Gottsdanker et al 1961; Schmerler 1976), so judgments of absolute distance and size from measurements like equations (12) and (13) may be crude. We found a solution for computing absolute distance and size, however, that does not require the measurement of retinal image acceleration. An observer is frequently able to see the apex of the flight of an object. In such cases, the solution is simpler because the initial velocity, V_y, of the object is zero at the apex, so the future velocity is simply gt.

Taking advantage of this observation, one can show that

$$Z = -\frac{gt(y-y_e)}{v_y y_e} . \qquad (14)$$

The solution here requires the measurement of only three retinal image properties: velocity, position, and the focus of expansion. In other words, one does not need to measure acceleration in the retinal image if the apex of the flight of the object is visible.

Equations (12), (13), and (14) show that one can in principle estimate absolute distance if the direction and accelerative effects of gravity are known. Both solutions involve measuring three or four different retinal image properties at an instant in time. Presumably other solutions are possible involving position or velocity measurements at different times, but we have not examined these.

Our analysis assumes that the observer knows the direction of gravity with respect to the line of sight. It seems reasonable that one could determine this angle from (i) a variety of environmental cues such as the apparent slant and tilt (in the retinal image) of environmentally vertical structures like walls, buildings, and trees, and (ii) vestibular and proprioceptive information that specifies gaze angle with respect to the vertical. The classic 'rod and frame' studies of Witkin and his colleagues illustrate the sensitivity of humans to both forms of cues in estimating vertical alignment of a stationary object (Witkin et al 1962). That work indicates that one might uncover interesting individual differences in the effective use of the gravity cue to depth and/or size.

Saxberg's (1987a) and the present analysis show that observers could judge the absolute distance and size of objects if they knew the accelerative effects of gravity (whether that is acceleration per se or distance expectancies for time from apogee). Saxberg (1987b) tested this proposition experimentally. He presented computer displays of balls propelled from below with different initial velocities. Observers were asked to move a mouse-driven plate on the computer screen to 'catch' the ball when it landed. Saxberg reported that observers were reasonably accurate when the ball provided information about both the trajectory and the changing retinal size. However, when the size of the ball was held constant on the screen, performance suffered significantly. From this, he concluded that human observers are not sensitive to the distance information contained in the dynamics of free fall.

As we see it, Saxberg's conclusion is not well-justified. There are at least three points of concern. First, the absolute distance of the catching plate was not specified; its distance was specified only by its position on a texture gradient and by the perspective changes in the shape of the plate as its distance was altered. These are relative depth cues. It is possible, however, that the observers were able to learn through practice the relationship between the position of the plate on the texture gradient and the specified absolute distance. Second, the observers viewed the displays binocularly, so they received information that the stimuli were all in the same depth plane. It is well known that binocular depth information can override or at least influence the interpretation of monocular depth cues. Third, and most important, Saxberg did not present a condition in which the use of information from free fall dynamics was presented in isolation. His conclusion was based on the decline in performance when the changing-size cue was removed but trajectory information was retained. Consider, however, the information the observer is provided with in the fixed-size condition. The trajectory information specifies appropriate changes in depth, but retinal size information specifies that no depth change has occurred. If observers are sensitive to changing retinal size information and trajectory information, this situation puts the two cues in conflict. In this situation, then, all that one can learn is whether observers are sensitive enough to trajectory information for this to override retinal size information.

Thus, in our view, Saxberg's (1987b) data do not provide persuasive evidence that human observers are insensitive to the absolute distance and the absolute size information contained in the dynamics of free fall. There is some indirect but persuasive evidence that observers in fact use such information. Cinematographers have long been aware of the problem of making miniature objects appear full-size when filmed. The problem arises when the miniature objects are set into motion. When a monster wrecks a skyscraper, the vertical acceleration ($\boldsymbol{a}_y$) across the film plane of the camera of the chunks of falling concrete is given by equation (11). For simplicity, we assume that V_x, V_y, and V_z are initially zero (corresponding to an object that falls directly downward from the apex of its flight or its initial resting position). Then

$$\boldsymbol{a}_y = \frac{g}{Z_0}, \qquad v_y = \frac{gt}{Z_0}, \qquad \text{and} \quad y = \frac{\frac{1}{2}gt^2}{Z_0}. \tag{15}$$

From these relationships, an observer can estimate Z_0 and perceive the small scale of the model. However, if the film speed is slowed down during projection, acceleration across the retina would signify a longer distance to the object and thus a larger object. How much should one slow the film speed? If a model is meant to give the impression of an object M times larger than its actual size and, of course, M times further away than its actual distance, then equation (15) suggests that one should create the impression that t has increased by a factor of M. That is to say, a model scaled down by a ratio 1:16 should be filmed at four times the normal speed. When the film is later projected at normal speed, it should give the intended impression of an object sixteen times larger than the model. Examination of the cinematography literature on special effects (eg Culhane 1981; Spottiswoode 1969) shows that this is exactly the rule-of-thumb that is used for projectiles and floating objects. At least one technical discussion relates the rule to 'gravity-fed motion'.

> Mobile miniatures are particularly difficult to work with owing to time-scale differences which exist between the model and the full-size original. If a moving miniature is to appear realistic, its speed must be decreased in proportion to its reduced scale, because all linear dimensions appear to be magnified as the square of the magnification of time. In the case of gravity-fed components, this reduction is achieved by over cranking the camera an appropriate amount. The formula employed here is $(D/d)^{1/2} = f$, where D is the distance or dimension in feet for the real object, d is the distance or dimension in feet for the miniature (this fraction being simply the reciprocal of the scale of the model) and f is the factor by which the operating speed of the camera is increased. (Spottiswoode 1969, page 727).

Other indirect evidence suggesting that observers are sensitive to the effects of gravity comes from an unpublished report by Johansson and Jansson (1967). They showed four films of divers diving into a pool. Subjects were asked to set the film speed to the value that made the event look natural. Settings were consistent within subjects and across films, varying from 5% to 15% of the projection speed. The authors did not report the constant errors, however, so we do not know if the observers exhibited biases in their settings. It follows from equation (15) that if time is scaled by a factor of N (eg the projector speed is increased or decreased N-fold), perceived size may be scaled by N^2. Consequently, relatively small deviations in projection speed can have significant (ie exponential) effects on perceived size. If subjects know the size of a typical diver and use their knowledge of the effects of gravity, one would expect them to adjust projection speed both consistently and accurately. Of course, other explanations for these specific data are possible.

Relative motion of limbs, head, and torso, as well as assumptions about inertia (mass) and average strength (force), could be combined to derive an estimate of appropriate film speed. Our preference for the gravity-cue hypothesis is that it is as equally relevant to nonbiological as to biological objects and it accounts for the special effects formulation.

3 Conclusion

We presented two solutions for estimating the absolute distance to a freely falling object. Equation (12) does not require that the apex of the flight be visible but does require a measurement of retinal acceleration. Equation (14) requires a visible apex but does not require retinal acceleration. Human observers are apparently not very good at discriminating object motions differing in acceleration (Gottsdanker et al 1961; Schmerler 1976), so we speculate that distance estimation is much better when the apex of the flight of an object is visible. (This may even provide a duplication of information when deceleration to the apex has been observed prior to acceleration from apex.) We assume that knowledge of absolute distance and retinal projection of even unfamiliar objects will allow direct evaluation of their absolute size. Again, however, we expect this estimation of size will be better when the apex of motion is visible.

As noted at various points in our discussion, knowledge of the absolute distance of an object will allow derivation of its absolute size and, likewise, knowledge of its absolute size will allow derivation of its absolute distance. This follows from the tangent rule in optical geometry as deftly applied by Berkeley long ago. It should be clear, therefore, that there are a number of options as to the specific manner in which gravity-influenced motion might be employed in object perception and motor adjustment to an object in free flight. One option, of course, is that it has no effect. Perhaps the necessary computations are too complex to be performed quickly enough to assist behavioral timing, for example. On the other hand, if the gravity cue is used, it may be that object distance is computed and this is followed by a secondary derivation (eg the tangent rule) of absolute size. Or conversely, perhaps size is computed directly and distance is derived in a secondary manner.

To summarize, we asked whether the dynamics of an object whose motion is influenced by gravity could provide information about the absolute distance to the object and the absolute size of the object. Formal analysis shows that, in principle, the dynamics do provide such information. Informal evidence from judgments of realism in filmed motion is consistent with the expectation that this available evidence about object distance and size is operative in human vision. We would guess, therefore, that when Newton saw the apple fall, he saw its size and distance.

References

Berkeley G, 1709 "A new theory of vision" in *Berkeley's Philosphical Writings* Ed. D M Armstrong (1965) (New York: Collier Books)

Chapman S, 1968 "Catching a baseball" *American Journal of Physics* **36** 868 - 870

Crowell J A, Banks M S, Royden C S, 1989, "A physiologically plausible model of optic flow perception" *Investigative Ophthalmology and Visual Science* **30** 427

Culhane J, 1981 *Special Effects in the Movies: How They Do It* (New York: Ballantine Books)

Gibson J J, Olum P, Rosenblatt F, 1955 "Parallax and perspective during aircraft landings" *American Journal of Psychology* **68** 372 - 385

Gottsdanker R M, Frick J W, Lockard R B, 1961 "Identifying the acceleration of visual targets" *British Journal of Psychology* **52** 31 - 42

Hofsten C von, 1990 "Catching" in *Tutorials on Perception and Actions* Eds H Heuer, A F Sanders (Hillsdale, NJ: Erlbaum)

Ittelson W H, 1951 "Size as a cue to distance: Static localization" *American Journal of Psychology* **53** 54 - 67

Johansson G, Jansson G, 1967 "The perception of free fall", Seminar delivered at the University of Uppsala (unpublished)
Lee D N, 1974 "Visual information during locomotion", in *Perception: Essays in Honor of James J Gibson* Eds R B McLeod, H L Pick (Ithaca, NY: Cornell University Press)
Lie I, 1965 "Convergence as a cue to perceived size and distance" *Scandinavian Journal of Psychology* **6** 109 - 116
Longuet-Higgins H C, Prazdny K, 1980 "The interpretation of a moving retinal image" *Proceedings of the Royal Society of London, Series B* **208** 385 - 397
Rogers B, Graham M, 1982 "Similarities between motion parallax and stereopsis in human depth perception" *Vision Research* **22** 261 - 270
Saxberg B V H, 1987a "Projected free fall trajectories: I. Theory and simulation" *Biological Cybernetics* **56** 159 - 175
Saxberg B V H, 1987b "Projected free fall trajectories: II. Human experiments" *Biological Cybernetics* **56** 177 - 184
Schmerler J, 1976 "The visual perception of accelerated motion" *Perception* **5** 167 - 185
Spottiswoode R, 1969 *The Focal Encyclopedia of Film and Television Techniques* (New York: Hastings House)
Todd J T, 1981 "Visual information about moving objects" *Journal of Experimental Psychology: Human Perception and Performance* **7** 795 - 810
Woodworth R S, Schlosberg H, 1954 *Experimental Psychology* (New York: Holt)
Witkin H A, Dyk R B, Patterson H F, Goodenough D R, Karp S A, 1962 *Psychological Differentiation: Studies of Development* (New York: John Wiley)
Yonas A, Pettersen L, Granrud C E, 1982 "Infants' sensitivity to familiar size as information for distance" *Child Development* **53** 1285 - 1290