

Using the free fall of objects under gravity for visual depth estimation

PIETER JAN STAPPERS and PATRICK E. WALLER
Delft University of Technology, Delft, The Netherlands

(Paul J. Locher, *Sponsor*)

We performed an experiment that tested people's ability to use the free fall of objects under gravity as a scale referent in a two-dimensional display. Subjects were presented with a two-dimensional computer animation of a fountain on a perspectival video background. Both rich and poor stimuli (i.e., fountains spraying many drops of water or only a single drop) were used. The task was to change the magnitude of simulated gravity so as to make the display appear "natural." Results indicate that subjects reliably matched gravitational acceleration to apparent depth. Accuracy and reliability of the settings were higher for the rich stimuli than the poor stimuli, demonstrating the importance of complex stimuli for the human perceptual system.

The present article reports the results of an experiment that tested whether people can use the free fall of objects under gravity to make accurate depth judgments in a video display. Estimating depth from a flat image is one of the traditional problems of perception. Images, both static and moving, are generally considered to be ambiguous with respect to the scale of the depicted objects because depth is lost in the formation of the image. As Eriksson (1974) has noted, "optical changes (irrespective of their complexity) are ambiguous with regard to absolute distance" (p. 225). To make absolute depth or size judgments, a scale referent such as the size of a known object is therefore needed. Most scale referents heretofore investigated have been static geometric sizes (e.g., object size or eye height; Sedgwick, 1973). A dynamic invariant, the vertical acceleration of objects in free fall, may also provide such a scale referent, as all falling objects undergo the same acceleration: $g = 9.8 \text{ m/sec}^2$ (insofar as air resistance can be ignored).

The experimental literature on visual motion perception suggests that people are very insensitive to visual accelerations (see, e.g., Calderone & Kaiser, 1989). One reason for this is that accelerations are assumed to be derived from the visual image by double differentiation, which is a very noise-sensitive operation (e.g., Koenderink, 1990). Most experimental results, however, have been obtained from "poor" stimulus conditions, which typically depict collisions in the picture plane, that is, isolated points accelerated in arbitrary directions and accelerated for a single moment in time. In contrast, the time scale of many complex movement patterns, such as an

animal walking or an object in free fall, is not arbitrary. Its rate is fixed by a constant, named the Froude number, that expresses an invariant relation between the magnitude of the gravitational acceleration constant and the spatial and temporal dimensions of dynamic processes (McMahon & Bonner, 1983).

The purpose of the present study was to show that, through this invariant, natural events provide a dynamic scale referent that observers may use to disambiguate depth in a perspective representation. In the present experiment, subjects were shown a video recording of a scene on which a computer-simulated fountain was superimposed. The subject's task was to make the display look "natural" by changing the magnitude of the gravitational acceleration used in the simulation. To do this, he/she had to match the magnitude of gravitational acceleration to the apparent depth.

Demonstrations like Johansson's (1973, 1975) point-light walkers suggest that our sensory system is better equipped to deal with complex natural motion patterns than with isolated dots. Furthermore, it was expected that subjects would be more accurate and reliable at matching acceleration to apparent depth for rich patterns, like fountains, than for traditional, poor patterns, like a single bouncing ping pong ball. To test this hypothesis, half of the experimental trials used a rich stimulus (a fountain spraying water drops), and the other half used a poor stimulus (a fountain with only a single drop of water visible at any time). The display consisted of a video background on which a two-dimensional computer animation of a fountain was shown. In each trial, a fountain appeared at a constant apparent depth, and the subject varied the magnitude of simulated gravity until the display looked natural; this required matching the simulated value of gravity to the apparent depth.

Correspondence should be sent to the first author at the Laboratory for Form Theory, Faculty of Industrial Design Engineering, Jaffalaan 9, NL-2628BX Delft, The Netherlands (e-mail: p.j.stappers@io.tudelft.nl).

Distances (m)

F 196
 E 154
 D 112
 C 70
 B 35
 A 17

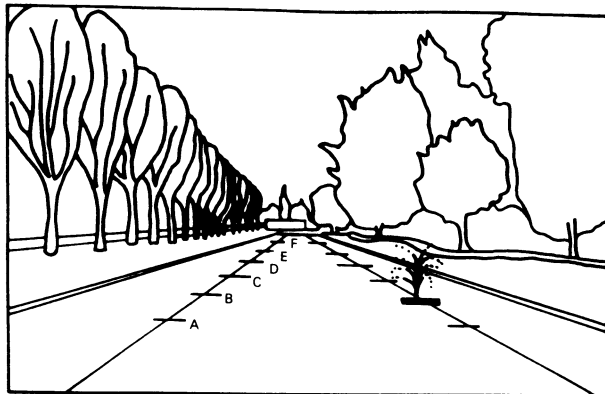


Figure 1. Outline drawing of the composition of the video background. The six positions (A-F) of the simulated fountain on the right and left side of the display are shown; depths for each of the six positions are shown above the figure.

METHOD

Design

The experiment used a repeated measures randomized block factorial design with two independent variables, apparent depth (six levels) and stimulus poverty (two levels). Apparent depth was manipulated by having the same animated fountain displayed in different positions on the screen, as shown in Figure 1. Different depths were used to eliminate possible bias of judgments in a video display. The depths were chosen to be almost equally spaced on the display, while covering a large range of apparent depths. At each depth, two fountain positions were used to eliminate right-left visual field biases.

Stimulus poverty was manipulated by the number of water drops used in the simulated fountain; a rich display contained 60 drops, a poor display only 1 drop. The dependent variable was the magnitude of gravitational acceleration as set by the subject, expressed in display units (pixels/sec²).

Stimuli

Stimuli consisted of a computer animation of a fountain shown on a perspectival video background of a canal in a park; it had compelling pictorial cues such as perspective (see Figure 1). The video recording displayed natural motions like people walking and ducks swimming in the background. Care was taken to ensure that no objects moved through display regions where the fountains were positioned. The animated fountain had a fixed size on the screen, but had a variable playback rate and position on the screen corresponding to variable values of gravity and apparent depth, respectively. The relation between display position and depth was measured in the scene. The number of drops of water was kept constant during the animation by "recycling" drops, when they hit the bottom level, at a randomized initial speed and direction.

The video background (played on a Panasonic AG-6200 VHS recorder) and the animation (generated in real time on an Archimedes A410 with ARM-3 processor) were combined using a Panasonic MX10 mixing panel and displayed on an AMIGA 1760 monitor. All subjects judged the resulting displays as convincing; the rich displays evoked a fountain, the poor displays a bouncing ping pong ball. The subjects controlled the value of gravity in the simulation by moving a mouse to left or right; the mouse's pointer was not shown on the screen.

Subjects

The subjects were 10 first- or second-year students of industrial design engineering who had little experience or training in perspective-drawing techniques.¹

Procedure

On each trial, the subjects adjusted the fountain animation so that the image on the screen looked as natural as possible. They took as long as they wished and indicated when they were satisfied with their determination by pressing the mouse's button.

The fountain appeared twice in each of the 12 positions in Figure 1. For one of the two trials for each position, the value of simulated gravity was initially set at a randomly selected magnitude that was 2-4 times larger than the physical value. For the second trial, the initial magnitude was 1/2-1/4 the physical value. Each subject completed 24 trials, which were randomized across subjects, during the experimental session, which lasted approximately 15 min.

RESULTS

Figure 2 shows average acceleration settings (responses) for the different conditions of apparent depth and stimulus poverty. An analysis of variance revealed significant main effects for stimulus poverty [$F(1,119) = 65.11, p < .001$] and apparent depth [$F(5,119) = 8.11, p < .005$] and no significant interaction [$F(5,119) = 2.04$].

Separate linear regression analyses were performed for poor and rich stimulus conditions to determine the relationship between acceleration and inverse depth. A high correlation was found for both the poor and rich conditions ($R_s = .87$ and $.86$, respectively). The regression lines are shown in Figure 2, together with the physical values for acceleration. The position and slope of each line shows that the subjects' settings varied with depth in the same way as the physical value. Moreover, the slope of the regression line for the rich condition is significantly ($p < .001$) closer to that of the physical value than is the slope for the poor condition (95% confidence intervals: rich, $B = 39 \pm 6$; poor, $B = 28 \pm 4$). Finally, the variance of responses in the rich condition was significantly lower than that in the poor condition [$F(119,119) = S^2 \text{ poor}/S^2 \text{ rich} = 1.94, p < .05$].

DISCUSSION

Our findings indicate that people are sensitive to the time scale of an environmentally likely display such as the one used in this study, namely, a stream of particles falling under gravitational acceleration. This kind of motion pattern, like Johansson's (1973, 1975) point-light walker displays, is common enough in the environment so that our perceptual systems may have become attuned to it in the course of evolution; it may constitute an invariant of ecological optics (Gibson, 1979). The fact that the subjects performed most accurately at the larger depths (which was also found in a pilot study) may be due to the temporal resolution (25 frames/sec) of the stimuli: the faster the animated drops move across the display, the "jumper" the display becomes. Another explanation may be that the range of human sensitivity to visual accelerations does not extend to gravitational free-fall events that occur very close to the observer.

The present experiment addresses sensitivity to the time scale of an ongoing event, but not the mechanism implementing it. One constraint by which the scale of the fountain is specified is through the instantaneous accelerations (a) of the drops, namely, $a = g$, where $g = 9.8 \text{ m/sec}^2$, the constant of gravity. Another constraint lies in the ratio between the vertical launching velocity (v) of a drop and the maximal height (h) it

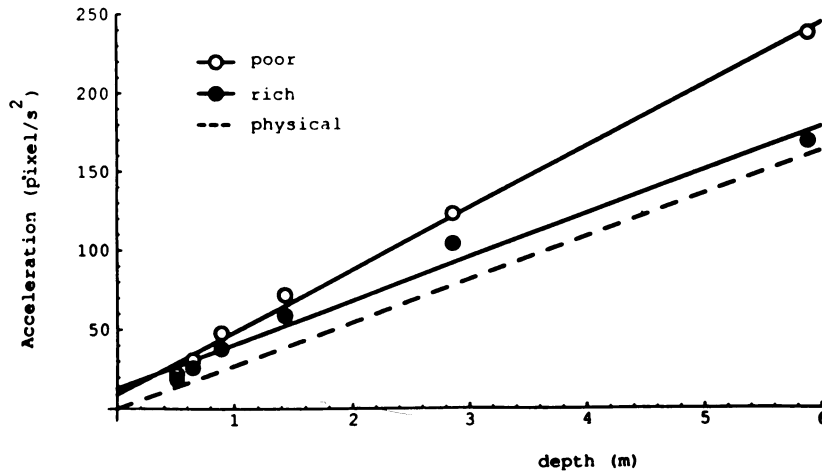


Figure 2. Gravity matching in screen units (pixels/sec²), averaged over all subjects. Depths are set out inversely along the horizontal axis because projected dimensions scale linearly with inverse depth. The dashed line indicates the physical value $g = 9.8 \text{ m/sec}^2$.

attains in its parabolic trajectory, because these are related by $h = v^2/2g$. The first constraint suggests double differentiation, the second a mechanism that operates over a time span rather than instantaneously. Although most computer vision research into perceptual mechanisms has concentrated on instantaneous (frame-by-frame) analysis, Gibson's (1979) arguments against the punctate conception of a stimulus in perception and Runeson's (1977) arguments on smart mechanisms point toward mechanisms operating over time.

The results of the present experiment indicate that for some natural events, optical changes are not, as Eriksson (1974) claimed, ambiguous with regard to absolute distance, because dynamic events are not ambiguous with regard to "absolute" time.

REFERENCES

- CALDERONE, J. B., & KAISER, M. K. (1989). Visual acceleration detection: Effect of sign and motion orientation. *Perception & Psychophysics*, **45**, 391-394.
- ERIKSSON, E. S. (1974). A theory of veridical space perception. *Scandinavian Journal of Psychology*, **15**, 225-235.
- GIBSON, J. J. (1979). *The ecological approach to perception*. Boston: Houghton Mifflin.
- JOHANSSON, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, **14**, 201-211.
- JOHANSSON, G. (1975). Visual motion perception. *Scientific American*, **204**, 48-52.
- KOENDERINK, J. J. (1990). Some theoretical aspects of optic flow. In R. Warren & A. H. Wertheim (Eds.), *Perception and control of self-motion* (pp. 53-68). Hillsdale, NJ: Erlbaum.
- MCMAHON, T. A., & BONNER, J. T. (1983). *On size and life*. New York: Scientific American Library.
- RUNESON, S. (1977). On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology*, **18**, 172-179.
- SEDGWICK, H. A. (1973). *The visible horizon: A potential source of visual information for the perception of size and distance*. Unpublished doctoral dissertation, Cornell University, Ithaca, New York.

NOTE

1. The subjects also took part in a second experiment in which the acceleration was constant and the subjects controlled the fountain's position in the display (matching apparent depth to acceleration). Variation in the responses in this experiment was extremely large, probably due to a methodological flaw, that is, as the subject adjusted the fountain's position, its apparent depth changed, but its height measured on the screen remained constant. This produced conflicting information, which severely impaired the realism of the display. Results from this experiment are not reported here.

(Manuscript received October 2, 1992.)