Matching perceived depth from disparity and from velocity: Modeling and psychophysics

Fulvio Domini\textsuperscript{a},*, Corrado Caudek \textsuperscript{b}

\textsuperscript{a}Brown University, Cognitive and Linguistic Sciences, Providence, RI 02912, USA
\textsuperscript{b}Università degli Studi di Firenze, Firenze, Italy

**ABSTRACT**

We asked observers to match in depth a disparity-only stimulus with a velocity-only stimulus. The observers’ responses revealed systematic biases: the two stimuli appeared to be matched in depth when they were produced by the projection of different distal depth extents. We discuss two alternative models of depth recovery that could account for these results. (1) Depth matches could be obtained by scaling the image signals by constants not specified by optical information, and (2) depth matches could be obtained by equating the stimuli in terms of their signal-to-noise ratios (see Domini & Caudek, 2009). We show that the systematic failures of shape constancy revealed by observers’ judgments are well accounted for by the hypothesis that the apparent depth of a stimulus is determined by the magnitude of the retinal signals relative to the uncertainty (i.e., internal noise) arising from the measurement of those signals.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

When a stationary object is viewed binocularly, the horizontal disparities created by the relief of the object can be used by the visual system to recover information about the three-dimensional (3D) shape. Likewise, when the same object is viewed monocularly, information about 3D shape can be gathered from the pattern of projected velocities generated by the relative motion between an observer and the object. In the present investigation, we will try to understand in which conditions these two different sources of depth information support the perception of an equivalent amount of depth.

1.1. Perceptual depth estimation from retinal images

It has long been recognized that disparity and velocity signals by themselves are insufficient to specify the distal depth map that has generated the two-dimensional (2D) image. For a local analysis of the visual field, in fact, the depth difference (\(z\)) with respect to the fixation point is related to the disparity (\(d\)) and velocity (\(v\)) signals through parameters which are not specified by optical information (see Fig. 1). For small visual angles, the equations become:

\[
d \approx \text{IOD} \frac{z}{z_f} + \epsilon_d,
\]

\[
v \approx \left(\frac{T_x \omega}{z_f^2}\right) \frac{z}{z_f^2} + \epsilon_v,
\]

where \(z\) is the depth-difference with respect to the fixation point, \(z_f\) is the fixation distance, IOD is the inter-ocular distance, \(T_x\) is the x-axis translation of the observer and \(\omega\) is the rotation of the object. \(\epsilon_d\) and \(\epsilon_v\) specify additive Gaussian noise with zero mean and standard deviations \(\sigma_d\) and \(\sigma_v\).

In order to compute the absolute amount of depth, the horizontal disparities must be scaled inversely with the square of the viewing distance, \(z_f^2\) (by assuming that the IOD is known). Likewise, the velocity signals must be scaled so as to take into account the observer’s translational component \(T_x\), the object’s rotation \(\omega\) and the inverse of the viewing distance. To simplify the following discussion, we will denote with:

\[
k_d = \frac{\text{IOD}}{z_f^2}.
\]
the scaling factor for veridical recovery of Euclidean depth from disparity information. Likewise, we will denote with:

\[ k_v = \frac{1}{z_d} \left( \frac{T_x}{\Delta z} + \omega \right), \]

the scaling factor for veridical recovery of Euclidean depth from velocity information.

In summary, the disparity and velocity signals are related to the depth of the projected object through parameters not specified by optical information. Disparity signals are related to depth through the squared reciprocal fixation distance \( z_f \); the velocity signals are related to depth through the reciprocal of the fixation distance, the angle of rotation \( \omega \), and the translation velocity \( T_x \). How can the “missing parameters” \( z_f \), \( \omega \), and \( T_x \) be recovered?

1.2. Perceived depth from disparity signals

The parameter \( z_f \) may be specified by not-retinal information, such as the vergence angle of the eyes and the state of the accommodation (e.g., Proffitt & Caudek, 2002). Some investigations have provided evidence that observers can estimate fixation distance from vergence (Frisby, Buckley, & Duke, 1996; Tresilian & Mon-Williams, 2000; Tresilian, Mon-Williams, & Kelly, 1999) and accommodation (Fisher & Ciuffreda, 1988). In general, however, extra-retinal information does not guarantee a veridical recovery of viewing distance. The vast majority of experiments on the perception of depth from binocular disparity, in fact, has shown systematic distortions of depth-from-stereo. In the context of the inverse-geometry models of depth perception, these distortions of depth have been attributed to a mis-estimation of the viewing distance (e.g., Johnston, 1991; for a discussion, see Todd & Norman, 2003).

Some of these experiments have been conducted in the laboratory with computer-generated displays (Bradshaw, Glennerster, & Rogers, 1996; Brenner & Landy, 1999; Brenner & van Damme, 1999; Collett, Schwarz, & Sobel, 1991; Glennerster, Rogers, & Bradshaw, 1996, 1998; Johnston, 1991; Johnston, Cumming, & Landy, 1994; Norman & Todd, 1998; Todd, Oomes, Koenderink, & Kappers, 2001; Tittle, Todd, Perotti, & Norman, 1995), whereas other studies have been carried out by using real objects in fully illuminated natural environments (Baird & Biersdorf, 1967; Battro, Netto, & Rozestraten, 1976; Bradshaw, Parton, & Glennerster, 2000; Cuijpers, Kappers, & Koenderink, 2000a; Cuijpers, Kappers, & Koenderink, 2000b; Gilinsky, 1951; Harway, 1963; Koenderink, van Doorn, Kappers, & Todd, 2002; Koenderink, van Doorn, & Lappin, 2000; Loomis, Da Silva, Fujita, & Fukusima, 1992; Loomis & Philbeck, 1999; Norman, Crabtree, Clayton, & Norman, 2005; Norman, Lappin, & Norman, 2000; Norman, Todd, Perotti, & Tittle, 1996 – see their Experiment 4). In most of the cases, however, the psychophysical literature suggests that human observers do not estimate the viewing distance correctly.

1.3. Perceived depth from velocity signals

The recovery of depth from velocity information, besides being undermined by the indeterminacy of \( z_f \), also suffers from the indeterminacy of the parameters \( \omega \) and \( T_x \). A veridical description of the Euclidean structure of an object can be derived from second-order optic flow, if appropriate assumptions are introduced in the interpretation process (Longuet-Higgins & Prazdny, 1980; Ullman, 1979). It has been shown, however, that human observers have a very limited sensitivity for the second-order temporal properties of the optic flow and thus rely mainly on velocities to recover 3D information (Hogervorst & Eagle, 2000; Todd & Bressan, 1990). If only the first-order information is used, however, it is not possible, in principle, to derive a veridical estimate of depth.

In our own research, we have showed that observers recover metric information from the optic flow in a heuristic and patchwork fashion. We provided evidence that this is achieved by a probabilistic process which assigns the most likely 3D interpretation to (ambiguous) local first-order properties of the optic flow. As a consequence, the perceptual interpretation of velocity information is, in general, neither veridical nor internally consistent (Caudek & Domini, 1998; Caudek & Proffitt, 1993; Caudek & Rubin, 2001; Di Luca, Domini, & Caudek, 2004, 2007; Domini & Braunstein, 1998; Domini & Caudek, 1999, 2003a, 2003b; Domini, Caudek, & Proffitt, 1997; Domini, Caudek, & Richman, 1998; Domini, Caudek, & Skirk-ko, 2003; Domini, Caudek, & Tassinari, 2006; Domini, Caudek, Turner, & Favretto, 1998; Domini, Vuong, & Caudek, 2002; Tassinari, Domini, & Caudek, 2008).

2. Perceptual depth-matching of single-cue stimuli

If the visual system were able to veridically estimate the “missing parameters” \( z_f \), \( \omega \), and \( T_x \), then two single-cue displays would elicit the perception of the same depth extent when they correspond to the projection of the same distal depth extent \( z \). The literature on the perceptual interpretation of disparity and velocity signals, however, reveals systematic failures of shape constancy (see the previous section). As a consequence, a disparity-only stimulus can be perceived as deeper or shallower than a velocity-only stimulus, even if the two stimuli are generated by the projection of the same distal depth extent, \( z_f = z_v \). Likewise, a disparity-only stimulus can be perceived as having the same depth extent of a velocity-only stimulus, even if the two stimuli are generated by the projections of two different depth extents, \( z_f \neq z_v \).

The goal of the present investigation is to explain these “depth mis-matches”. Two different hypotheses will be contrasted.

---

Fig. 1. Schematic illustration of the viewing geometry. \( z_f \) is the viewing distance; \( z_d \) and \( z_v \) are the angles formed by the visual lines connecting a flanking point to the left and right eyes and the two visual axes; \( \Delta z \) is the front-to-back depth of the stimulus configuration along the line of sight. \( \omega \) represents the angular rotation about a vertical axis centered at the fixation point. The open circles represent the position of the flanking points after the rotation. \( z_f \) and \( z_d \) are the angles between the visual line connecting a flanking point to the cyclopean eye and the visual axis of the cyclopean eye before and after the rotation, respectively.
2.1. Hypothesis 1: Perceived depth is recovered by estimating the “missing” scaling constants

According to the first hypothesis, perceived depth is estimated by reversing the following equations:

\[ d = k_0 z + \epsilon_d, \]
\[ v = k_0 z + \epsilon_v, \]
where \( k_0 \) and \( k_v \) are defined as indicated by Eqs. (3) and (4). Misperceptions of 3D shape are attributed to the mis-estimation of the scaling factors \( k_0 \) and \( k_v \). For example, stereo-processing may fail to appropriately scale horizontal disparities because of the noise in the measurement of eye positions, or because of a conflict between the focus cues and the vergence angle, or because vertical disparities are not sufficiently reliable. Similarly, motion-processing may not accurately measure the head translation \( T_h \) or, in the case of a stationary observer, may not accurately estimate the amount \( \omega \) of object rotation.

The limitation of this account is that we cannot determine a priori whether two single-cue stimuli will be perceived to have the same depth or not. Before running the experiment, in fact, we have no way of establishing whether the estimates of the scaling factors \( k_0 \) and \( k_v \) are accurate or biased.

2.2. Hypothesis 2: Perceived depth depends on the magnitude of the image signal relative to the uncertainty

Depth perception from disparity or velocity signals does not necessarily require the estimation of the scaling factors \( k_0 \) and \( k_v \), as defined by Eqs. (3) and (4), respectively. Recently, we have advanced a different proposal, according to which the apparent depth of a stimulus is determined by the magnitude of motion and/or disparity differences in a display relative to the uncertainty (i.e., internal noise) arising from the measurement of those signals (Domini et al., 2006; Tassinari et al., 2008). Our proposal is formulated in terms of a normative model of depth-cue integration (the intrinsic constraint model), but it can also be applied to depth recovery from single cues and to depth-matching tasks. According to the intrinsic constraint model, depth estimates are recovered through a weighted combination of image signals. This combination is optimal in the sense that it maximizes the signal-to-noise ratio (SNR) of the combined estimate. A thorough discussion of the intrinsic constraint model is provided by Domini and Caudef (2009). What is important for the present discussion is that, according to the intrinsic constraint model, two stimuli elicit the same amount of perceived depth when they exhibit the same ratio between signal intensity and discrimination threshold. According to such proposal, therefore, a disparity-only stimulus will be perceived with the same depth as a velocity-only stimulus when they both provide the same signal-to-noise ratio. Note that, in such “retinal model,” the properties of the distal scene are not considered. In the following section, we will illustrate how such proposal can be applied to the stimulus conditions of the present investigation.

2.2.1. Relation between discrimination thresholds and perceived depth magnitudes

The disparity and velocity signals are proportional to the relative depth \( z \) and, thus, directly specify the affine structure of the distal 3D shape (see Eqs. (5) and (6)). We hypothesize that a perceptual depth match occurs whenever two image signals specify the same affine structure with the same degree of precision, that is, whenever two image signals have the same SNR:

\[ \frac{\epsilon(d)}{\sigma_d} = \frac{\epsilon(v)}{\sigma_v}. \]

This hypothesis allows us to predict the disparity and velocity values of a perceptual depth match. In the present experiment, a staircase procedure was used to match the perceived depth of a varying disparity-only stimulus to the perceived depth of a fixed motion-only stimulus. According to Eq. (7), a depth match requires the expected value of the disparity signal to be

\[ \frac{\epsilon(d)}{\sigma_d} = \frac{\epsilon(v)}{\sigma_v}, \]

where \( \epsilon_0 \) is the front-to-back relative-velocity of the velocity-only stimulus.

An estimate of \( \epsilon(d) \) is provided by the point of subjective equality (PSE) of the psychometric function estimated from the observers’ judgments. Such point of subjective equality (termed PSE) was obtained in the first part of the experiment, where the disparity-only stimuli were varied within a staircase procedure and the velocity-only stimuli were kept fixed at the value corresponding to the signal intensity \( v_0 \).

In the second and third part of the experiment, we measured the discrimination thresholds JND \(_d \) and JND \(_v \) (expressed in terms of the signal’s intensities) at the pedestal values \( PSE_d \) and \( v_0 \). Such discrimination thresholds are interpreted as estimates of \( \sigma_d \) and \( \sigma_v \), respectively. We can therefore re-write Eq. (8) as

\[ PSE_d = \frac{\text{JND}_d}{\text{JND}_v} v_0. \]

Eq. (9) can be re-written in terms of the depth of the distal 3D shapes used to generate the stimulus displays. According to Eqs. (5) and (6), \( PSE_d = k_0 PSE_z, \quad v_0 = k_0 z_r, \quad \text{JND}_d = k_0 \text{JND}_z, \) and \( \text{JND}_v = k_0 \text{JND}_z \). By substituting these values in Eq. (9), after simplifying for \( k_0 \) and \( k_v \), we obtain:

\[ PSE_z = \frac{\text{JND}_d}{\text{JND}_v} z_r. \]

In conclusion, the systematic errors made by observers in matching the perceived depth of two single-cue stimuli can be predicted by Eq. (10) in terms of the uncertainty (i.e., the JND) of the perceptual responses. According to Hypothesis 1, conversely, the biases in the perceptual matching tasks cannot be predicted, since there is no way to determine by how much the scaling constants \( k_0 \) and \( k_v \) are mis-estimated.

3. Experiment

The experiment had a twofold purpose: (1) to determine the stimulus conditions in which a disparity-only stimulus is perceived to be matched in depth to a velocity-only stimulus, and (2) to measure the discrimination thresholds of the depth-matched stimuli.

3.1. Apparatus

Stereoscopic stimuli were displayed on a haperscope consisting of two CRT monitors (0.22 mm dot pitch) located on swing arms pivoting directly beneath the observer’s eyes. Anti-aliasing and spatial calibrating procedures allowed spatial precision of dot location greater than hyper-acuity levels. Each monitor was seen in a mirror by one eye. Head position was fixed with a chin-and-forehead locating apparatus. The distance from each eye to the corresponding monitor was 95 cm. The eyes’ vergence was directly manipulated by physically moving the monitors on their swing arms. Since the monitors and mirrors pivot rigidly about the eye’s axis of rotation, the retinal images always remain the same for all
positions of the two CRT monitors (see Fig. 2). Thus, the changes in eye position were dissociated from the changes in retinal images.

### 3.2. Stimuli

The stimuli were 800 high-luminance anti-aliased dots displayed against a low luminance background. 400 dots were randomly placed on three vertical lines 50 mm long (see Fig. 4). The other points were randomly positioned within a volume 50 mm wide, 50 mm high and 25 mm deep.

The actual physical distance to the CRT monitors (suggested by accommodation cues) was always fixed at 95 cm from the observer. In the experiment, we manipulated the eyes' vergence so as to specify two possible viewing distances: 50 and 100 cm.

In different trials, the center line and the two flanking lines were spaced in depth by either 2.5 or 5.0 mm. One of the three vertical lines was positioned at the center of the stimulus display. The other two vertical lines were positioned at 12.5 mm to the left and to the right of the central line. The overall stimulus subtended about 2.86° of visual angle when viewed at 0.5 m; when viewed at 1.0 m, the visual angle was 1.43°.

Depth information was provided by either disparity or velocity cues. Stimuli were drawn in polar projection. Disparities were calculated so as to simulate a 3D structure viewed at 50 or 100 cm from the observer. The vergence angle was computed for each observer, by taking into account her or his inter-ocular distance. For both stereo and motion stimuli, the simulated depth position of the two flanking lines was at fixation.

In the motion-only condition, observers viewed the stimuli with both eyes. This allowed us to equate the vergence information provided in both motion-only and disparity-only conditions. The same image was provided to both eyes, thus producing a nil disparity field. In our previous research, we found that a nil disparity field does not flatten depth from motion for small simulated depth magnitudes.

The 2D motion of the dots in the display was computed by simulating a rotation of the simulated 3D structure about a horizontal axis positioned at fixation. The 3D structure rotated back and forth by 14°. The duration of an entire rotation cycle was 2 s. The rotation increment was 0.23° per frame. The update rate was 60 frames per second.

### 3.3. Procedure

Observers were asked to determine which of two successively presented stimuli evoked a larger depth separation. The experiment comprised three parts.
of the experiment. In a third part of the experiment, the JNDs for the velocity-only stimuli were estimated by means of a depth-discrimination task at the simulated values of 2.5 and 5.0 mm used in the first part of the experiment (see Fig. 3).

3.4. Participants

A total of three undergraduate and three graduate students from Brown University participated in the experiment. They all had normal or corrected-to-normal vision. All participants were naïve to the purpose of the study, and three of them were experienced psychophysical observers. Participation was voluntary; all participants provided written consent and the undergraduate students were paid for their participation.

3.5. Results and discussion

For each of the three parts of the experiment, the data of each observer in each experimental condition were fitted with a psychometric function. Fig. 5 shows the data of one representative observer for the first part of the experiment.

The means of the fits are points of subjective equality, the values of the velocity-only stimuli that on average had the same apparent depth as the stereo stimuli. If perceived depth were veridical, the PSEs of the psychometric functions would be the following: 123.759 arcsec for the blue function; 30.940 arcsec for the red function; 247.518 arcsec for the green function; 61.880 arcsec for the black function. For this particular observer, judgments are close to veridical at the viewing distance of 1.0 m; at the viewing distance of 0.5 m, instead, the responses of R.F. over-estimated depth from disparity.
This pattern of results is confirmed by Fig. 6, which shows the average results of six observers expressed in terms of depth magnitudes (not in terms of signal intensities, as in Fig. 5). In the left panel of Fig. 6, the average depth magnitudes of the disparity-only stimuli at the PSE are shown as a function of the simulated depth of the velocity-only stimuli. The viewing distances of 0.5 and 1.0 m are as coded red circles and blue triangles, respectively. Note that, if the depth matches were veridical, all points should lay on a straight line with zero intercept and unitary slope (the gray line in the Fig. 6).

The effect of viewing distance was confirmed by a linear mixed-effects (LMEs) model applied on the PSEs with two factors as independent variables: simulated depth (2.5, 5.0 mm) and viewing distance (0.5, 1.0 m). The interaction between simulated depth and viewing distance was not significant (χ² = 2.488, n.s.). A no-interaction model revealed significant effects for both simulated depth (β = 0.611, SE B = 0.105, MCMC p = 0.0002) and viewing distance (β = 0.003, SE B = 0.0005, MCMC p = 0.0001).

Separate analyses were then performed at each of the two viewing distances. For the viewing distance of 1.0 m, we found that the depth of the disparity-only stimuli at the PSE did not differ from the simulated depth of the velocity-only stimuli (β = −0.688, SE B = 0.520, MCMC p = n.s.). At the viewing distance of 0.5 m, the depth of the disparity-only stimuli at the PSE was significantly lower than the simulated depth of the velocity-only stimuli (β = −1.968, SE B = 0.383, MCMC p = 0.0002). In summary, the present data suggest a systematic bias: a perceptual depth match at the viewing distance of 0.5 m requires about half of the simulated depth for a disparity-only stimulus than for a velocity-only stimulus.

Having established that, within the present stimulus conditions, the perceptual depth matches are biased, the purpose of the present investigation was to determine whether such biases can be accounted for by the hypothesis described in Section 2.2. According to Hypothesis 2, a perceptual depth match requires the same signal-to-noise ratio for the two single-cue stimuli, regardless of their simulated depth magnitudes. At the PSE the depth magnitudes of the two stimuli should be related to each other as indicated by Eq. (10). The mean PSE of the disparity-only stimuli as a function of the predictor of Eq. (10) is shown in the right panel of Fig. 6. The data of the individual observers are shown in Figs. 7 and 8.

Hypothesis 2 can be tested by estimating the 95% confidence interval for the slope of this empirical regression line and to gauge it against the model’s predictions.

---

1 Inferential statistics are based on a linear mixed-effects (LMEs) model specifying participants as random effects (Pinheiro & Bates, 2000; Quené & Van den Bergh, 2004, for simulations). We used the lme4 program (lme4 package; Bates & Sarkar, 2007) in the R system for statistical computing (R Development Core Team, 2009). As indicated by Baayen (2008), p values and confidence intervals are generated from the posterior distribution of parameter estimates with Markov Chain Monte Carlo methods, using the mcmcsamp program in the lme4 package with default specifications (e.g., n = 1,000 samples; locally uniform priors for fixed effects; locally non-informative priors for random effects).

2 One outlier data point was deleted from Fig. 8. The data analysis was performed on the entire data set.
4. General discussion

In the present investigation, observers were asked to discriminate the amounts of depth perceived from disparity-only and from velocity-only stimuli. Given the psychophysical evidence discussed in Sections 1.2 and 1.3, it would be naïve to expect that, when they are both generated by the same distal depth extent, observers judge a velocity-only stimulus to be as deep as a disparity-only stimulus. In general, perceived depth from single-cue stimuli is distorted. The purpose of the present investigation was to understand in which conditions the coupleings of different distal depth extents support the perception of a depth match.

According to the first hypothesis presented in Section 2, a disparity-only display may appear to be matched in depth to a velocity-only display simulating a different depth magnitude because the scaling constants $k_d$ and $k_v$ are mis-estimated (see Eqs. (3) and (4)). Since we do not know the amount of departure of the $k_d$ and $k_v$ estimates from their true values, we cannot predict a priori when disparity-only and velocity-only stimuli will be perceived as having the same depth.

According to the second hypothesis, two single-cue stimuli are perceived to be equally deep when they both provide the same SNR. This hypothesis stems from the idea that the amount of perceived depth depends on the retinal SNR. Consistent with this hypothesis are, for example, the findings of Todd (1985) who studied structure-from-motion by manipulating the amount of visual noise in the displays. In his study, observers were required to estimate the apparent slant of a random-dot planar surface rotating in depth, on which a proportion of the dots were repositioned randomly on each frame. The results revealed that the apparent slants of the surfaces decreased systematically as the proportion of the displaced dots was increased. If external noise can reduce the magnitude of apparent depth or slant, then it seems reasonable to suppose that internal noise may have the same effect.

More specifically, we proposed that a perceptual depth match is affected by the uncertainty deriving from the measurement of the image signals, as indicated by Eq. (10). It follows that, in general, two single-cue stimuli should be perceived as matched in depth when they are the projection of specific different distal depth extents. The results of the present investigation are consistent with this hypothesis.

At the viewing distance of 0.5 m, a perceptual match was obtained when the distal 3D shape of which the disparity-only stimulus was the projection was 49% shallower than the distal shape for the velocity-only stimulus (see Fig. 6). At the viewing distance of 1.0 m, the distal depth for the disparity-only stimulus was 16% shallower than the distal depth of the velocity-only stimulus (a not significant difference). In other words, a perceptual depth match was obtained, at the viewing distance of 0.5 m, when the two stimuli were generated by the projection of two very different distal depths magnitudes: at the viewing distance of 1.0 m, conversely, the distal depths for the two stimuli were quite similar.

As indicated in the right panel of Fig. 6, these data are consistent with Hypothesis 2. In all cases, in fact, the two stimuli had to provide the same SNR in order to appear to be matched in depth. Note that the predictions of Eq. (10) hold both when the perceptual matches are “veridical” and when they are not.

Alternative explanations are also important to consider. It may be argued, for example, that observers can interpret correctly disparity information at the “effective viewing distance” of 100 cm, but they over-estimate depth from disparity at the distance of 50 cm (see Johnston, 1991). Disparity information coupled with a vergence angle specifying the abathic distance, together with a specific mis-estimation of the angle of rotation, thus, could be held responsible for the results of the present experiment. This alternative explanation, however, misses an important aspect of the present phenomenon. In fact, even if depth from disparity were always over-estimated at small viewing distances and under-estimated at large viewing distances, what about depth from velocity? From our previous research on the perceptual interpretation of the optic flow, we know that apparent depth from motion depends not only on the distal depth magnitude, but also on the amount of angular rotation used to generate the velocity field (for a review, see Domini & Caudé, 2003a). Any explanation of the present results only in terms of the role of the abathic distance in disparity processing, therefore, is incomplete.

Another consideration can be made about the fact that our data set is too small to convincingly show that the slope of the relationship shown in Fig. 6 is 1.0. This point can be addressed by applying Eq. (10) to a larger data set provided by the investigation of Domini and Caudé (2009). In that paper, we investigated the relation between two Fechnerian depth scales. One psychophysical depth scale for velocity-only stimuli, the other for disparity-only stimuli. In the third part of that experiment, observers were asked to perceptually match a disparity-only stimulus with a velocity-only stimulus. In a similar manner as in the present experiment, observers judged which of two successively presented stimuli evoked a larger depth separation. The simulated depth separation of the velocity-only stimulus was kept constant, whereas the depth separation of the disparity-only stimulus was varied according to a staircase procedure. The PSE for the disparity-only stimulus was found, together with the JNDs of both stimuli at the PSE.
Fig. 9. Left panel. Depth (in millimeters) of disparity-only stimuli as a function of the depth (in millimeters) of the velocity-only stimuli, when the two stimuli are perceived as perceptually matched in depth. The data have been replotted from Domini and Caudenk (2009). Vertical bars represent one standard error. Right panel. The same data of the left panel. The values on the horizontal axis have been recoded according to Eq. (10).

In Fig. 9 we have replotted the data of Domini and Caudenk (2009) in the format of Fig. 6. Each point in the plot represents the distal depth magnitudes used to generate a disparity-only and velocity-only stimuli at the PSE. Note that systematic failures of shape constancy occur also in the experiment of Domini and Caudenk (2009). The left panel of Fig. 9 shows that the amount of distal depth for the disparity-only stimuli is consistently smaller than the distal depth for the velocity-only stimuli.

At the PSE, the distal depth magnitudes of the disparity-only stimuli are approximately linearly related to the distal depth magnitudes of the velocity-only stimuli (Fig. 9, left panel). The slope of this linear relation, however, is different from one (gray line, with a nil intercept) and represents the systematic depth misperceptions in depth-matched stimuli. In the right panel of Fig. 9, the PSEs of the disparity-only stimuli are shown as a function of the simulated depths of the velocity-only stimuli weighted by the ratio between the JNDs of the two stimuli. The gray line in the right panel of Fig. 9 has nil intercept and slope of 1.0. It is easy to see, therefore, that Eq. (10) provides an accurate account also for the data of Domini and Caudenk (2009). Note an important difference between the stimulus conditions of Domini and Caudenk (2009) and those of the present investigation. Differently from the present experiment, in fact, Domini and Caudenk (2009) always kept the viewing distance fixed at 1.0 m. The depth distortions found in that investigation, therefore, cannot be attributed to a mis-estimation of the viewing distance.

Another interesting result of the present investigation is the effect of the vergence angle on the perceptual interpretation of velocity information. At the viewing distance of 0.5 m, the simulated depth of 2.5 mm generated a maximum displacement of 50.4 arcsec; at the PSE, the matched disparity-only display exhibited a maximum disparity of 66.7 arcsec. At the viewing distance of 1.0 m, the simulated depth of 5.0 mm generated a maximum velocity of 50.4 arcsec; at the PSE, the matched disparity-only display exhibited a maximum disparity of 53.8 arcsec. The observers’ responses in these two conditions were not statistically different ($t_s = 1.255, n.s.$). This means that the same retinal displacements (50.4 arcsec), when coupled with different fixation distances (0.5 and 1.0 m), were perceptually matched with (almost) the same disparity (60.25 arcsec, on average) viewed at different distances. Even though there may be a real difference between these two conditions, which was not detected by the statistical power of the present analysis, this result suggests a potential interesting issue. We know that the same disparity signal coupled with a larger viewing distance produces an increase in the amount of perceived depth (Johnston, 1991). The present data, therefore, are compatible with the hypothesis that the vergence angle may affect the perceptual interpretation of velocity information: perceived depth may increase when the same velocity field is associated with a larger fixation distance. Even though it is consistent with Eq. (2), the possible effect of the vergence angle on the perceptual interpretation of the velocity field has not been previously addressed in the structure-from-motion literature.

A final point concerns the conflict between vergence and accommodation. Even though the conflict between these two depth-cues may hinder “veridical perception”, it is not a problem for the present investigation. In fact, it affects the visual processing of disparity and velocity information in the same manner.

5. Conclusion

Two single-cue stimuli are not necessarily perceived to be matched in depth when they are generated by the projection of the same distal depth magnitude. In the present investigation, we showed that a disparity-only stimulus elicits the same amount of depth as a velocity-only stimulus when they both provide the same signal-to-noise ratio, regardless of the distal depth magnitudes that have generated the two images. Such empirical results are consistent with the hypothesis that the apparent depth of a stimulus is determined by the magnitude of the retinal signals relative to the uncertainty (i.e., internal noise) arising from the measurement of those signals (Domini & Caudenk, 2009; Domini et al., 2006).

References


For the other two points of Fig. 6 the 2D signals are the following. For the smallest simulated depth of the motion displays at the distance of 1.0 m, the displacement signal is 25 arcsec with a matched disparity of 28 arcsec; for the largest simulated depth of the motion displays at the distance of 0.5 m, the displacement signal is 100.8 arcsec with a matched disparity of 128 arcsec. In this case it is difficult to interpret the effect of the vergence angle on the processing of velocity information because a “small” disparity paired with a “large” fixation distance must be compared with a “large” disparity paired with a small fixation distance. In these circumstances, any prediction requires the knowledge of how the vergence angle scales disparity information. Such a function would be known if the fixation distance were estimated correctly by the visual system. But this is not so.