Journal of Experimental Psychology: Human Perception and Performance

A New View Through Alberti's Window

Maarten W. A. Wijntjes Online First Publication, February 3, 2014. http://dx.doi.org/10.1037/a0035396

CITATION

Wijntjes, M. W. A. (2014, February 3). A New View Through Alberti's Window. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. http://dx.doi.org/10.1037/a0035396

A New View Through Alberti's Window

Maarten W. A. Wijntjes Delft University of Technology

In his famous treatise on perspective, Alberti compared picture perception with looking through a window. Although Alberti himself was more concerned with picture production than perception, the window metaphor is still widely used to describe picture perception. By performing depth perception experiments, we investigated whether Alberti's hypothesis makes sense in a geometrical fashion. If pictures are regarded as windows, the locus of objects with equal depth should be similar for pictorial and real space-ideally, spherical. Furthermore, if the loci of equidistance are indeed similar for real and pictorial space, their difference should be flat. We designed two experiments to investigate this claim. In the first experiment, a pairwise depth comparison task was used to compute the global perceived depth structure of a complex scene. We found that perception of the real space is more accurate and less ambiguous than pictorial space. More interestingly, we found that the relative differences between these two spaces (locus of relative equidistance) are curved, which contradicts the window hypothesis. In the second experiment, we wanted to measure the absolute locus of equidistance that we believed was diagnostic for the difference between real and pictorial space perception. We found that under normal circumstances, the distribution of equally perceived depths is curved in real space, and relatively flat in pictorial space. However, we also found exceptions. For example, viewing real space with one eye yielded similar results as normal pictorial space perception. We conclude that Alberti's hypothesis needs a revision.

Keywords: visual perception, depth perception, pictorial space, stereo vision, space perception

The relation between pictures and reality has been intriguing art historians, philosophers, psychologists, and, inevitably, artists themselves. The fundamental difference between pictures and reality is that there is only one version of reality, while there are many ways to depict it. Some styles, like those used in ancient Egypt, are highly symbolic and bare little resemblance with (optical) reality, whereas others, such as the Pompeian murals (and later Renaissance art), are more optically oriented. Both optical and symbolic styles can be found throughout art history and are omnipresent in our contemporary daily life. The question arises whether there is an underlying mechanism that explains the perception of both these pictorial extremes. Two simple solutions would be that *all* pictorial styles are visually processed as being either symbolic (e.g., Kepes, 1995) or optical (e.g., Gibson, 1954). Gibson (1971) himself came to disprove both his own optical hypothesis as well as the symbolic hypothesis. He reasoned that a picture can be visually interpreted because pictures contain the same optical information for an observer as reality does; the invariants of visual perception are present in both pictures and reality. The validity of this view depends on how "information" and "invariance" are understood. The invariance of a face may be the spatial configuration of facial features. Depictions of faces

often comply with this invariance, but there are famous exceptions—such as Picasso paintings—that do not prevent us from perceiving the strokes of paint as faces.

For an empirical account on the difference between pictures and reality, we need to operationalize the question, that is, frame it into an experimental task. There are many qualities of a percept that are worth investigating and that can be used to describe the difference between reality and pictures. In this study, we consider depth perception. In reality, we always perceive depth, and in pictures, although they are physically flat, we generally perceive depth depending on the depicted scene. In both reality and in pictures, depth is a quality in the sense that it is not directly available to the brain. Depth is added to the two-dimensional retinal signal. It could be that the process of conscious depth formation is different when viewing reality as opposed to a picture.

There is not much literature that addresses this issue. Direct comparisons between real and pictorial space are rare. For example, Hecht, van Doorn, and Koenderink (1999) measured perceived angles in construction works in reality and in photographs. They found a compression of space with respect to veridicality, but similar results between reality and photographs. These, and other findings made Cutting (2003), in a review on this topic, conclude that "perceiving photographic space [is] like perceiving environmental space" (p. 216). A substantial gap in the current literature on visual space perception is that there are no data that describe the global depth structure. Most studies focus on single points in either real or pictorial space but do not analyze across the various loci of the visual field. Haber (1985) indicated that an exception on this locally oriented research comes from the imagery literature, in which Multi-Dimensional Scaling (MDS) techniques are used to quantify the global structure of perceived layout (e.g., Kosslyn,

The author was supported by a VENI grant from the Dutch Science Foundation (NWO).

Correspondence concerning this article should be addressed to Maarten W. A. Wijntjes, Perceptual Intelligence Lab Industrial Design Engineering, Delft University of Technology, Landbergstraat, 2628CE, The Netherlands. E-mail: m.w.a.wijntjes@tudelft.nl

Pick, & Fariello, 1974). However, these studies focus on the view-independent layout (top view) of a scene, which is related to, but different from, visual depth perception. In a typical MDS analysis, the distances between locations (or stimuli) serve as input. These pairwise distances can only be positive. However, for depth perception, the pairwise depths can be either positive or negative, depending on which object is in front.

The geometry of a visual scene can be split up in *shape* and *space*. 3D shape has been studied thoroughly—it is concerned with pictorial relief of single objects (e.g., Koenderink, van Doorn, & Kappers, 1992; Todd, 2004). Reconstructing the pictorial relief from individual attitude settings is essentially a global approach. In contrast to 3D shape, the spatial relations between objects in a visual scene have only recently been investigated in a global fashion. In these studies, pairwise depth comparisons were measured by either a pointing task (Wagemans, Van Doorn, & Koenderink, 2011a; Wijntjes & Pont, 2010), a relative size task (Wagemans, Van Doorn, Koenderink, 2011b), or a depth discrimination task (Van Doorn, Koenderink, & Wagemans, 2011). In the reconstruction algorithms that were developed alongside these new experimental tasks, the question emerged whether to include the

eye of the observer in the geometry. Wagemans et al. (2011a) used an orthographic projection system to reconstruct depth, but Wijntjes and Pont (2010) used perspective, in which the virtual observer was positioned at the center of projection. The latter method essentially treads the picture as a "window through which I see what I want to paint" (Alberti, 1435/1970, emphasis added). The geometry of this Albertian inspired picture-as-a-window hypothesis is illustrated in Figure 1A. The orthographic method used by Wagemans et al. (2011a) neglects the position of the observer. This makes sense in a formal way, because the observer cannot physically enter the (virtual) pictorial space. Presence of the observer is a major difference between real and pictorial space. Although not physically present, if the observer regards a picture as a window, he may be present in a geometrical fashion that is indistinguishable from real space. His presence can be detected by measuring the shape of the locus of equidistance, a concept we introduce here. The locus of equidistance is the surface of all points that have equal depth. This concept is comparable with the horopter but has nothing to do with stereo disparities, only with perceived depth. If the shape of the locus of equidistance is curved (ideally, spherical), then we could say that the observer is "pres-



Figure 1. Graphical representation of the hypotheses under investigation. (A) The geometry that is implied by Alberti's window hypothesis. (B) The geometry of an absent observer. (C) Subtracting two spaces of identical structure results in a *flat* locus of relative equidistance. (D) Subtracting a hypothetical spherical real space from a hypothetical flat pictorial space results in a *curved* locus of relative equidistance.

ent," his (virtual) location being the center of the curvature as Figure 1A. This finding would support that picture perception follows Alberti's window hypothesis. However, if the locus of equidistance is flat, we could argue that the observer is absent (or very far away), as illustrated in Figure 1B. We investigated which of these two hypotheses best describes the difference between real and pictorial space. We deliberately use the terms "real space" and "pictorial space" to accentuate the possibility that these spaces may be categorically distinct, that is, having different geometries (Koenderink & van Doorn, 2008). An alternative to this categorization is to speak explicitly in terms of informational differences (depth cues) between the two viewing modes. Visual space would be a "full cue" stimulus, whereas pictorial space would be a reduced cue stimulus. This would make sense if we a priori have computational reasons that the informational differences (like stereo information) result in the distinction between an Albertian viewing mode and an "absent observer" mode as shown in Figures 1A and B. However, to our knowledge, there are no a priori reasons for this prediction. Therefore, for now, we will maintain our real and pictorial space distinction.

Returning to our overview of recently developed depth perception methods (Van Doorn et al., 2011; Wagemans et al., 2011a, 2011b; Wijntjes & Pont, 2010): All studies use pairwise depth estimates that are integrated into a global depth structure. Furthermore, all studies used a probe (either a 3D arrow or a simple disk) that was shown in pictorial space, on a computer screen. It is easy to see that some of these tasks are more difficult to implement into a "reality" experiment than others. The pairwise depth comparison task used in the first experiment of this research only requires the rendering of two static dots in the scene, which is fairly easy to work out technically. With this method we are able to do basic performance analysis (like interobserver similarity and veridicality), but it can also be used to quantify the shape of the *relative* equidistance locus. To directly measure the equidistant locus, one would ideally adjust the distance of all objects in a scene until they appear at equal distance. This is generally not preferable because it destroys the structure of the scene. If we want to keep the structure of the scene intact and want to measure the difference between different presentation modes (e.g., pictorial or real), we can measure the relative equidistant locus by analyzing the difference between the two depth structures. By subtracting two depth structures, we essentially nullify the individual depth differences between the objects. This procedure is illustrated in Figure 1C and D. If the relative equidistant locus is flat (Figure 1C), then pictorial space is similar to real space and Alberti's hypothesis holds. If, on the other hand, the relative equidistant locus is curved (Figure 1C), then real and pictorial space are different, which could indicate that the pictorial locus of equidistance is flat, refuting Alberti's hypothesis.

The reason that we a priori doubt the picture-as-a-window hypothesis is that a picture is, in many respects, different from a window. What we see through a window often drastically changes when we change our viewpoint, whereas a picture is relatively invariant with respect to viewpoint changes. Furthermore, candidate images that, in principle, could serve as a window when the observer is in the correct viewing position are only those that are drawn or shot using linear perspective. Artists and graphic designers have invented many more drawings systems (White, 1987; Willats, 1997) that do not comply to linear perspective rules but nevertheless effectively depict depth. Thus, in daily life, we see many different styles of pictorial spaces under a large variety of viewing conditions. It seems likely that the human visual system does not have specialized strategies for each of these spaces and viewing circumstances, but rather has a general strategy that may be rooted in visual perception of reality but avoids computational problems associated with Alberti's window. One alternative could be "the-observer-is-absent," as depicted in Figure 1B.

In the first experiment, we used the pairwise depth comparison method to quantify the difference between real and pictorial space. The analysis of the relative equidistance loci gave considerable evidence that there is indeed a systematic difference between reality and virtuality. To investigate whether this effect was robust, we designed a second experiment in which we used a rather different scene. We also used a different method, one with which we could directly probe the shape of the absolute equidistance locus.

Experiment 1

In the first experiment, we used the pairwise depth comparison task to quantify several geometric properties of depth perception. We quantified the similarity of perceived depth orders, the internal consistency, and the veridicality. We also analyzed the relative equidistance loci.

Method

Participants. In total, 40 observers participated in this experiment. Ten observers performed the experiment in three conditions; the other observers participated in a single condition. This will be clarified in the Experimental Design section. All observers had normal or corrected-to-normal vision.

Stimulus and apparatus. The stimulus was an installation located in a lab space. It was built by the author, who collected various objects and positioned them with the intention to create a complex 3D scene. The installation was photographed with a Canon 5D Mark II (sensor size: 36 mm \times 24 mm) using a focal distance of 24 mm. Therefore, the perspectively correct ratio of viewing distance divided by screen width is two thirds, that is, a horizontal visual angle of 74°. The pictures were shot at F14 to minimize defocus effects. A stereo version was shot using a camera slider and taking a second picture that was shifted 6.5 cm horizontally, perpendicular to the optical axis. The optical axes of the stereo pair were parallel.

Initially, distances of 110 locations in the scene were measured using a DISTO 500 laser distance meter. From this set, a (computerized) random selection was made to optimize homogeneity in the picture plane within a relatively small distance interval. The final set consisted of 20 sample points (see Figure 2) that varied between 2.512 m and 3.663 m in distance.

The stimulus was viewed either in reality or in pictorial space. The pictorial space conditions were tested on two screen sizes. The small screen was a Wheatstone stereo setup with two LaCie (Electron22BlueIV) CRT monitors. Resolution for this screen was set at 1600×1200 pixels. Viewing distance was 66 cm and screen width was 40 cm, resulting in a 34° horizontal field of view. This angle is about twice as small as the perspectively correct viewing angle. Observers were presented with either a similar pair of

Figure 2. The spatial installation. The 20 sampling points are depicted by the dots.

pictures (the binocular condition) or a stereoscopic pair (the stereoscopic condition). The other screen was a Panasonic plasma TV (TX-P65VT30) on which the stimulus could be presented with a width of 129 cm. Viewing distance with respect to this screen was 86 cm, resulting in the perspectively correct horizontal field of view of 74°. This screen was either viewed with one eye (the monocular condition) or with two eyes (the binocular condition). Note that, similar to the viewing conditions for the small screen, the term binocular refers to viewing with two eyes, without stereoscopic information.

Experimental design. All conditions and observer groups are shown in Table 1. The 10 observers of Group A participated (all in fixed order) in the binocular small screen, stereoscopic small screen, and real condition, respectively. The other groups only participated in a single condition. Group D was essentially similar to the last condition of Group A and was used to check if the results of Group A were influenced by some kind of experience/ learning effect. The condition names are found in Table 1 but will also be referred to by the group name, with Conditions A1 to A3 referring to the three conditions performed by Group A. For example Condition A3 refers to the "real stereo" condition.

The choice of the various conditions deserves some clarification. A basic comparison between real and pictorial space is given by Conditions C and D. They are the most common and natural circumstances under which humans view real and pictorial spaces. A major difference between these conditions is stereo information. In Condition D, the stereo information correlates with the 3D geometry of the scene, whereas in Group C, the stereo information correlates with the flatness of the screen. Possible differences could be caused by either of these stereo cues, which is the rationale behind the other conditions. In Condition B, observers viewed the screen with one eye, which effectively removes the flatness information of the screen. This could result in observers being less aware of the screen and thus resembled the real condition more than the binocular condition. Because we were unable to run our code of the pictorial stereo presentation on the large screen, we had to revert to an existing Wheatstone setup with conventional small computer screens, which is the rationale behind Conditions A1 and A2. The reason behind Condition A3 is that we initially wanted a repeated measures design, but later on decided that we also wanted different groups of observers for the different conditions.

Thus, the basic reason behind the "extra" Conditions A1 through A3 and Condition B is to reveal the effect of stereo information. That we now also have a small-screen condition is essentially "collateral damage," but as we will see later, it reveals an interesting effect. The reason why we could not experiment with a real condition in which the viewing angle was similar to the small screen is lack of space in the experimentation room. The correct viewing distance in the real condition for the 34° viewing angle would be $\frac{\tan(74^\circ)}{\tan(34^\circ)} \approx 5$ times larger, that is, the average distance of

about 3 m would have to be increased to about 15 m.

Procedure. In all conditions, observers were shown two locations in the scene and asked which location appeared closer. All possible pairs of the 20 sample locations were measured, resulting in 190 judgments. In the screen conditions, two dots initially blinked to attract attention. The observer acknowledged that he was aware of the two dots by a pressing a key, after which the blinking stopped and static dots were shown. Then, the observer used the mouse to select the closest point. The (circular) dots were 7 pixels in diameter. The dots in Figure 2 are about 3.5 times larger, for clarity of presentation.

In the real conditions, observers were seated in front of the installation with their cyclopean eye at the same location from which the photo was taken. Observers were led into the room. They were not explicitly encouraged or discouraged to view the experimental setup, but the setup was visible. The experimenter walked them to the chair. They came from behind the original camera position, so they did not approach the installation closer than their designated viewing position. Viewing angles were not limited to the installation: Observers could also see the surrounding. The experimenter stood behind the observer and controlled two laser pointers that were attached to tripods. On a laptop screen (invisible to the observer), the experimenter could see where to point the lasers. The observers responded verbally concerning which laser (red or green) pointed to the closest location, which

Table 1

Overview of Participant Groups and Experimental Conditions

Group A	Group A	Group A	Group B	Group C	Group D
Pictorial small binocular	Pictorial small stereo	Real stereo	Large screen monocular	Large screen binocular	Real stereo

Note. The icons are used to denote the conditions in subsequent figures.



was registered by the experimenter. All experimentation software was written with PsychToolbox (Brainard, 1997; Pelli, 1997).

Data analysis. From the pairwise depth comparisons, an overall depth order was computed. It has previously been proposed that counting votes (the closest of the pair gets a vote) is a robust method to compute the best fitting depth order (Van Doorn et al., 2011). In the Appendix, we show that using a least-squares method approach yields the same result. Whereas we use this method to quantify depth perception, it is potentially useful for other pairwise psychophysical tasks, such as material perception (which is "glossier," instead of which is "closer").

To compute the internal consistency of a single session, the responses are first transformed into an overall depth order. This depth order is used as input to construct an "ideal" response set that follows from that specific depth order. Because the depth order algorithm allows for points to have the same depth order value (when they have an equal amount of votes), the "ideal" response set may contain "undecided" answers that do not occur in the actual response set. Therefore, these undecided responses were omitted. For the remaining responses, the ratio of consistent answers was computed. We performed numerical simulations to quantify the chance level for the consistency measure. Random depth structures (N = 10,000) were generated together with a set of random answers. This simulates an observer who is completely guessing. The consistency computed in this way has a chance level of 0.64666 (close, but not exactly two thirds).

Next, the depth orders were compared between observers of the same condition by computing Kendall rank-order correlation coefficients. High correlations imply high depth perception similarities between observers. Differently stated, the intersubjective ambiguity is low for high correlations, and vice versa. To quantify how well observers performed with respect to veridicality, the veridical distances were transformed in depth orders, which were correlated with the perceived depth orders.

To quantify the relative equidistance locus between viewing conditions, the data were averaged over subjects per viewing condition. The difference data Δz_i (see Figure 3 for an explanation) of each pair of viewing conditions were analyzed. Three types of models were fitted to the data: zeroth order (straight), first order (affine), and second order (curved). The straight model is simply $\Delta z = a_0$ (1). The affine model is $\Delta z = a_0 + a_1 x + a_2 y$ (2). It has been shown that the affine model can account for a significant amount of variability in depth differences in studies on pictorial

real

relief (e.g., Koenderink, van Doorn, Kappers, & Todd, 2001). The second-order model can account for nonlinearities. To understand the contributions of the horizontal and vertical directions, the second-order model was split in three models: horizontally curved $\Delta z = a_0 + a_1 x + a_2 y + a_3 x^2$ (3); vertically curved $\Delta z = a_0 + a_1 x + a_2 y + a_3 x^2$ (4); and doubly curved $\Delta z = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2$ (5). An example of the last model fitted to the data is shown in Figure 3.

To select the best of these five models, the Akaike information selection criterion was used (Wagenmakers & Farrell, 2004). This procedure selects the best-explaining model, taking into account the number of parameters. For example, if the affine model is selected, the differences between viewing conditions can best be explained by affine differences as opposed to either zeroth or second-order terms. The actual (absolute) values of the best-fitting parameters are meaningless, as we only address depth *order*. However, the sign of the best fitting parameters is an important indicator for the *qualitative* difference between viewing conditions. Especially, the sign of the nonlinear coefficients are of interest because they reveal whether the relative equidistance locus is convex or concave. The signs of the (significantly relevant) parameters will be graphically presented in the Results section.

Results

The internal consistency is plotted in Figure 4A. For Group A (see Table 1), the repeated measures ANOVA indicated a significant main effect, F(2, 18) = 16.819, p < 0.05, $\eta^2 = 0.651$. Post hoc pairwise comparisons (Bonferroni corrected) showed that consistency was lower in the binocular condition compared with both the stereoscopic and real conditions. No difference was found between these latter two conditions. For Groups B-C-D the (between observers) ANOVA also showed a significant main effect, F(2, 27) = 11.389, p < 0.05, $\eta^2 = 0.457$. Here, post hoc pairwise comparisons showed that in the real condition. Overall, these results reveal that when stereoscopic information is available (the stereoscopic and real condition), global depth perception is more consistent than the monocular and binocular conditions without stereoscopic information.

To assess depth order similarities between observers, Kendall rank correlations were computed within conditions. In each condition, 10 observer data sets were pair-wise correlated, resulting in

difference



stereo





Figure 4. Overview of the results in box plot format. The white lines denote medians, the boxes mark the first and third quartile, and the whiskers denote the lowest and highest data points. Significant (p < 0.05) Bonferroni-corrected post hoc tests are indicated by asterisks. (A) The internal consistency defined by the fraction of answers consistent with the overall depth order. Conditions are denoted by icons that are explained in Table 1. (B) Kendall rank correlation coefficients between observers in the same viewing condition. (C) Kendall rank correlation coefficients between observers and veridicality.

45 pairs. The repeated measures ANOVA for Group A revealed a significant main effect, F(2, 88) = 17.022, p < 0.05, $\eta^2 = 0.279$, as did the ANOVA for Groups B-C-D, F(2, 132) = 32.433, p < 0.05, $\eta^2 = 0.329$. The result is presented in Figure 4B. Overall, the results are similar to those found in the internal consistency analysis, except for the stereo condition. Although, no significant difference was found for the internal consistencies, the depth orders correlated better in the real condition than in the stereo condition. Nevertheless, the stereo condition correlates better than the mono condition.

Furthermore, depth orders were compared with veridicality. The actual distances were measured with an accurate laser distance meter. These distances were converted into depth orders. A significant main effect was found for Group A, F(2, 18) = 4.890, p < 0.05, $\eta^2 = 0.352$, as well as for Groups B-C-D, F(2, 27) = 6.267, p < 0.05, $\eta^2 = 0.317$. In Figure 4C, it can be seen that viewing the real scene results in better depth perception than viewing the pictorial scene. Only for the stereoscopic condition, the results are undecided—it is neither better nor worse than the monocular or real condition. Viewing the box plot of the stereoscopic condition, one is tempted to infer that the stereoscopic condition is equal to the monocular condition and fails to be statistically lower than the real condition because of the high variance.

Relative equidistance loci. Computing relative equidistance loci between all viewing condition pairs including the veridical depth order gives a total of 21 pairs. Figure 5 explains how to interpret the graphically presented models. For clarity, the condition comparisons were subdivided by defining three groups: two pictorial conditions (small and big screen) and the real conditions, as shown in Figure 6.

A quick glance at Figure 6 shows that, in almost all cases, one of the second-order models (see Data Analysis for an explanation) could best explain the relative equidistance loci. Cases in which the zeroth or first-order model were selected are very specific: always within a small-screen, big-screen, or reality condition, as indicated by the dashed square. For the reality condition, this is hardly surprising, but for the stereoscopic and binocular smallscreen conditions, this is perhaps a surprising finding.

When comparing the small-screen conditions with all other conditions (Figure 6A), a positive horizontal curvature was present in all cases. The vertical curvature was found to be less consistent. The binocular condition showed a negative curvature with respect to both large-screen conditions. All other comparisons showed a zero (flat) vertical curvature, except the comparison with veridicality that showed a positive curvature.

The large-screen conditions were compared with the remaining conditions (Figure 6B). In this case as well, the horizontal curvature was consistently positive (with one exception) and the vertical curvature was only present when compared with veridicality. Lastly, the real conditions (Figure 6C) did not show a positive horizontal curvature with respect to veridicality, although there was a positive vertical curvature present.

Curvature effects in the vertical direction are mainly absent, which is possibly caused by the narrower sampling in that direction (see Figure 2). A positive curvature is only present when compared with veridicality. In contrast, the horizontal direction results are very systematic. The way the data are presented in Figure 6 reveals that there is a certain hierarchy of the curvature effects with respect to the experimental conditions. To understand



Figure 5. An explanation of the graphically presented results of the model selection procedure that is presented in Figure 6. The medium gray horizontal line and the dark gray vertical line together represent the model. In this example, the relative equidistance locus between the binocular big screen and reality condition was modeled. The horizontal (medium gray) line is curved and the vertical (dark gray) line is straight, but slanted. This combination implies that the Akaike model selection criterion was best for Model 3 (see Data Analysis section of Experiment 1), the horizontally curved model. The "sign" of the horizontal line (i.e., concave) indicates that objects on the left and right sides are perceived to be closer in reality than in pictorial space. The backward slanted vertical line means that objects higher up in the visual field appear farther in reality than they do on a screen.

what the curvatures represent, consider that the theoretical equidistance locus is spherical. It follows that a positively curved difference between two conditions implies that the second condition is relatively less curved. This means that the equidistance locus for observers viewing in pictorial space is less curved than real space. Furthermore, the small screen elicits less positive curvature than the big screen. These findings indicate that there is a certain hierarchy of (qualitative) curvature differences regressing from a spherical locus in reality to a more planar equidistance locus in the pictorial spaces, depending on screen size. A larger screen resembles reality better than a small screen. This is visualized in Figure 7.

Discussion

The first finding that deserves attention may appear trivial but is nevertheless relevant. All data show a sufficient amount of noise to be useful for the analysis but, at the same time, are by far not completely random. In a pilot experiment that preceded the experiment presented here, we found that if sampling was too "easy," all responses were perfect and ceiling effects prevented performing any kind of analysis. If a metric method is used, such as a relative size task, there is always a certain amount of variance that can be analyzed. But in an ordinal method, chances are rather high to encounter zero variance if the task is not sufficiently difficult.

The results presented in Figure 4 reveal an interesting role for the disparity cue. It is to be expected that depth perception becomes better when more information is available (Cutting & Vishton, 1995). But how do we define "better"? If "better" is understood as being more veridical, then the results presented in Figure 4C cannot confirm this hypothesis. However, in terms of a higher internal consistency (Figure 4A) and a higher similarity between observers (Figure 4B) in depth orders, the disparity cue indeed makes depth perception "better." Similar findings have been reported in a study in which the pointing task was used to compare space perception in monocular and stereoscopic images (Wijntjes & Pont, 2010). In that study, the consistency was not found to be significantly improved by disparity information (although the overall trend was in that direction), but interobserver correlations did improve for two out of three stereo image stimuli. Furthermore, viewing from the perspectively correct point of view did not seem to increase any of the three statistics (consistency, interobserver correlation, and veridicality) in comparison with the incorrect viewing positions of the small-screen conditions.

The curved loci of relative equidistance are the most novel and perhaps most surprising findings of the first experiment. The window hypothesis implies that the space of the viewer and the pictorial space are essentially similar, merely separated by a window. This would mean that the equidistant loci are similar for real and pictorial space (a flat relative equidistance locus), which is clearly not in line with the current findings. Instead, a hierarchy of equidistance loci has been found that asks for a revision of Alberti's window. A hypothesis diametrically opposed to Alberti's window is that "the eye is not in pictorial space" (Koenderink & van Doorn, 2008; Wagemans et al., 2011b), or as we put it in the instruction, "the-observer-is-absent." In reality, the observer is located in the same space as the perceived objects. This location is probably used to relate distances to. In the case of looking at images, the observer is in a real space, whereas the objects are in a pictorial space. Hence, "the eye is not in pictorial space." This formulation clearly describes the inherent difference between real and pictorial space, and we can use it to conceive three hypotheses concerning how the observer might solve this problem. First, the observer may simply use his physical location with respect to the screen and infer depths from there. This is basically Alberti's hypothesis. Second, he uses an imagined location with respect to the screen. This may be useful when he is in the incorrect viewing



Figure 6. Geometric differences between all conditions, presented in table form. The elements denote the depth orders in the left column minus the depths orders in the top row. The dark gray vertical line shows the qualitative fit in the y-direction and the medium gray horizontal line in the x-direction. The bold font denotes the condition group that was compared with the other condition groups (regular fonts). The dashed frames denote comparisons within screen size/reality conditions.

position and can correct for that by imagining himself in the correct viewing position. Looking from perspectively incorrect viewpoints without experiencing any severe perceptual consequences is sometimes called La Gournerie's paradox. Literature on this topic is mainly focused on single shapes (e.g., Cutting, 1987; Vishwanath, Girshick, & Banks, 2005) and does not seem to have

solved the problem completely. The implications for the global depth structure have not been investigated. This second hypothesis of an imagined (instead of physical) viewpoint is still Albertilike—it regards the picture as a window. The third hypothesis is that the observer uses a view-independent strategy. In this case, objects with similar depths are organized in flat planes that are



Figure 7. Visualization of the hierarchical structure of the equidistant loci. Note that the data only show this trend in the horizontal direction.

parallel to the picture plane (possibly up to an affine transformation, but certainly no nonlinear transformations). No physical of imagined location of the "eye" is involved here—it is only pictorial space. To investigate this issue further, we designed a second experiment. We wanted to see if the (absolute) locus of equidistance is flat in pictorial space (as suggested by the third hypothesis) and investigate the effect of viewing the picture from the incorrect viewing position.

Experiment 2

The goal of this experiment was to get direct access to the shape of the absolute locus of equidistance, instead of relative loci of equidistance, between viewing conditions. To increase the generalizability of our findings, we also wanted to use a different kind of scene. To directly measure the shape of the locus of equidistance, we built a dense scene of hand-sized objects, distributed on a table. We designated one object in the middle as the reference object and asked the observer, for all other objects, whether they were closer of further away than the reference object. This results in a segmentation defined by the locus of equidistance.

For practical reasons, we designed a configuration of objects that predominantly varied in the horizontal plane. Therefore, we also only analyzed the curvature in the x-direction. Furthermore, we added another condition: oblique viewing. In the real world, a rotation of the viewer should result in a similar rotation of the locus of equidistance. However, if the viewer changes its position with respect to the screen, the scene content does not rotate along with the viewer. We reasoned that if the observers regards a picture as a window, that the locus of equidistance should rotate along with the viewer position.

Method

Participants. A total of 30 observes participated in this study, five per condition. All had normal or corrected-to-normal vision.

Stimulus and apparatus. The stimulus was a collection of small (approximately hand-sized) objects placed in a table. The table was covered with photo studio background paper. The objects were quasi-randomly placed such that objects were generally not completely occluded by other objects. The stimulus is shown in Figure 8. As can be seen in the top view, the standard object (the carrot) to which all other objects were compared is located in the middle. The photo was taken with the same camera as Experiment

1, using similar focal distance settings (24 mm). In the virtual conditions, observers viewed the similar screen as used in Experiment 1. However, the settings were a little different. Due to so-called overscan settings, a TV, by default, uses only part of the available pixels to display a computer image. During the preparation of the second experiment, we changed these settings, which resulted in a total image width of 144 cm, instead of 129 cm used in Experiment 1. Observer distance was scaled accordingly, that is, to 94 cm instead of the 86 cm in the first experiment. In both the real and pictorial conditions, observers used a chin rest. The veridical distances were again measured with a laser distance meter. Because the instruction in this experiment was to compare the average positions of the objects (this turned out to be more intuitive than the surface locations on which the laser pointed), we measured the average positions. This was done by pointing the laser to a small target surface that we held above the middle of all objects.

Experimental design. Observers performed the experiment in either of six viewing conditions of a 2×3 design. The first factor was real or pictorial presentation. The second factor was monocular orthogonal, binocular orthogonal, and binocular oblique. The first two of these conditions are basically similar to those used in Experiment 1, although in this experiment we added the monocular real condition. Binocular in the pictorial condition implies viewing the monocular picture with two eyes, whereas in the real condition, this implies stereo information. The new condition in the second experiment was oblique viewing. In this condition, the observer position was shifted 20° to the left with respect to the setup. Distance between the observer and the standard object (the carrot) was kept constant. It should be noted that the picture was kept similar, that is, no new picture was taken from the oblique viewpoint. This means that in the pictorial condition, the observers were not in the perspectively correct viewpoint.

Procedure. A curtain hung in front of the stimulus so that no observer saw the actual setup before the start of the experiment. In the real condition, observers were led to the designated position. After the instructions and after the observers had put their chin in the chin rest, the curtain was removed. Observers were instructed "to say whether these [indicated] objects are further or closer away from you than the carrot." Sixty-two objects were subsequently pointed at in random order by a laser pointer in the real conditions and a red dot in the pictorial conditions.



Data analysis. To quantify the shape of the locus of equidistance, we used a second-order polynomial embedded in a logistic function, which we will shortly explain. A parabola is normally written as $y = a_2x^2 + a_1x + a_0$, but can also be written as the solution of f(x, y) = 0 with $f(x, y) = a_2x^2 + a_1x + a_0 - y$. Note that "y" has the same meaning here as "z" in Experiment 1. The parabola can be used to model the locus of equidistance and we define that f(x, y) > 0 means further away, and f(x, y) < 0 means closer. As an example, this can be implemented in a Heaviside step function with H(f[x, y]) = 0 for f(x, y) < 0 and H(f[x, y]) = 1 for $f(x, y) \ge 0$. In principle, this function can be used to model ing. Therefore, we used a logistic function that uses an extra parameter σ that denotes the slope of the transition. The modeling function can be written as

$$g(x, y) = \frac{1}{1 + e^{(a_2 x^2 + a_1 x + a_0 - y)\sigma}}$$
(1)

where g is the (mean) response. Examples with different parameters are shown in Figure 9.

Results

First, we will analyze the average data, and second, we will address individual differences. The average data together with the best model fit are shown in Figure 10. If we first focus on the nonoblique conditions, it can be seen that the binocular pictorial and monocular real condition both show a relatively flat equidistance locus, whereas the monocular pictorial and stereo real condition are more curved. The locus of equidistance for the stereo real condition are most curved. Furthermore, the oblique conditions show a rotation in the same direction as the change of viewing position, although this is more evident in the real condition than in the pictorial condition. As will become clear from the individual data, there was one observer in the oblique pictorial condition who was responsible for the rotation of the pictorial data.

Parameters of the model are presented in Figure 11. The error bars denote confidence intervals attained from the nonlinear fitting procedure. These parameter values show that the curvatures of the binocular and oblique pictorial conditions, together with the monocular real conditions, are close to zero. The nonzero curvature of the monocular screen condition is considerably lower than the stereo real condition (a factor of 3) and oblique real condition (a factor of 2). Furthermore, the rotations (parameter a_1) are all close to zero, except for the oblique real condition.

The individual data show the same trend as the average data but also reveal substantial individual differences. From the average data, it appeared that the curvature found in the monocular pictorial condition was substantially larger than the binocular pictorial condition (Figure 11, top). But in the individual data, this difference appears to vanish. Moreover, the stereoscopic pictorial condition shows a rather large variability. The difference in the real condition between monocular and stereoscopic is similar, as found in the average data.

Rotation has little effect on the pictorial condition, except for one observer (first in row in Figure 12A). However, the rotation had a marked effect in the real condition, similar to that found in the average data.

Discussion

At first sight, the data give a clear picture. The average data are mostly in line with the individual data. However, the individual differences are prominent. Besides the oblique condition, there was one condition that we did not use in the first experiment, which was monocular viewing of the real scene. For completeness, we had incorporated the monocular condition this time. In the first experiment, we investigated the effect of stereoscopic information in the small-screen conditions and we did not find differences concerning the shape of the relative equidistance locus. Therefore, we did not expect many differences in the real condition. Surprisingly, we did find a rather large difference between the monocular real condition is of comparable magnitude as those found in the pictorial conditions. This implies that closing one eye changes depth perception from a reality-like into a pictorial-like structure.

The results in the oblique conditions are in line with both the second (mental view point correction) and third (view independent strategy) hypotheses we formulated in the end of the discussion of Experiment 1. Having an oblique (i.e., perspectively wrong) point of view in the pictorial condition yields similar results as the



Figure 9. Visualization of the logistic function with the embedded parabola. The upright axis denotes the response g that is modeled by (1). (A) Prediction of pictorial condition: the locus of equidistance is a flat plane. (B) Straight but sloped; this could happen in the real oblique condition. (C) Prediction of the real condition: Observers perceive objects of similar distance around a curved (possibly circular) surface. (D) Prediction of the oblique real condition; both a curvature and a rotation due to the changing viewpoint.



denotes the average response, and the background rendering denotes the best fitting model. The width of the gradient transition denotes the slope of the logistic function.

orthogonal viewing condition, except for one observer. In the real condition, the locus of equidistance rotates with the rotated point of view. This reflects that, in reality, the eye is indeed in the same space as the objects.

Lastly, it was surprising that about one fifth of the participants asked the experimenter (the author) how they should interpret the



Figure 11. Best-fitting parameters including confidence intervals from the fitting procedure. Top: curvature parameter a_2 . Bottom: rotation parameter a_1 .

instructions. In the first experiment, only two or three observers asked the experimenter for clarification. In both cases, they were not given any additional instructions and were told that they could interpret it in whatever way they wanted, but were encouraged to use the same interpretation throughout the experiment. We got the impression that some observers were indeed asking whether they should use a flat equidistance measure or a curved one (in our terminology). This possible cognitive penetration of their perceptual behavior is beyond the scope of the current article but should certainly not go unnoticed.

General Discussion

This study tried to reveal whether there are any geometric differences in the perception of real and pictorial space. There are many a priori informational differences between these two viewing modes. The long, but likely incomplete, list of informational differences is pixel resolution (limited in pictorial condition); disparity resolution (thresholds are likely smaller than pixel sizes); head movements (we used a chin rest in the second experiment, but this does not completely cancel out motion parallax); framing (in reality, observers saw more of the surrounding than in the pictorial condition); accommodation and vergence cues (not present in pictorial condition); color (screen has a limited color gamut); and dynamic range (screen has a limited dynamic range). All of these differences may or may not be of influence to our current findings. We believe that before investigating the contribution of various informational resources, it is important to study the generic case. For most people, the daily visual experience is a mixture between real space and pictorial spaces such as paintings, posters, and



Figure 12. (A) Individual data visualization including the best fitting models. (B) Individual curvature parameters a_2 . Note that because we used five observers per condition, the whiskers, box sides, and red line all (by definition) indicate one observer. (C) Individual rotation parameter a_1 .

electronic displays. In experimental psychology, the motivation behind using pictures is mostly practical. They are used as a surrogate for the real-life situation (such as the object recognition literature). To make results of pictorial experiments generalizable, it is of vital importance to understand the informational differences between the two viewing modes. Our study differs from this because we are not regarding pictorial space as a practical surrogate but as an equal ecologically valid viewing condition as reality itself. The main focus of this study was to understand if the two generic cases are differently perceived, and while doing so, we chose to additionally investigate only the most prominent informational difference between reality and pictorial space: stereoscopic information.

In comparing real and pictorial space, the influence of stereoscopic information may a priori be twofold. Actual disparity information is present in reality and may be recreated in pictorial space by using a pair of stereo images. This would imply that stereoscopic pictorial space would be more like real space than monocular/binocular pictorial space. And looking with one eye to a real scene would make reality more pictorial. The second possible effect is that when looking with two eyes at a monocular picture, stereo information tells the observer that he is looking at a flat display. This implies that looking with two eyes accentuates the pictorialness of pictorial space, whereas viewing with one eye may result a more "reality" kind of experience. In the first experiment, we found that stereoscopic information has a positive effect on basic perceptual performance (see Figure 4) in the stereo pictorial condition. But we did not find any performance difference between looking with one or two eyes at the same picture. In the locus of equidistance analysis, practically no differences were found both in the mono/stereo comparison of the small-screen conditions and the mono/bino comparison of the large-screen conditions. It could be that differences were simply too small to be revealed by our analysis, as we did find them in the second experiment. The curvature parameter of the average data in the monocular pictorial condition are in the direction of those found in stereoscopic real space, although this effect seems to vanish in the individual data analysis. More prominently, looking at reality with one eye dramatically changes the locus of equidistance to be as flat as the pictorial conditions. This is in line with the commonly used trick by artists to close one eye when painting. This may not have been known to Alberti, but later in history, Da Vinci was quite clear about this: "Objects in relief, when seen from a short distance with one eye, look like a perfect picture" (Da Vinci, 2004).

Our findings indicate that Alberti's hypothesis does not hold. Perception of spatial structure in pictures is not similar as

12

looking through a window (Alberti, 1435/1970; Gombrich, 1960). It should be noted that what Alberti meant with his window comparison is heavily disputed (Masheck, 1991). It was probably intended to be a functional account on how linear perspective works and not originally targeted to explain the psychology of picture perception. Nevertheless, it is a useful concept to build a discussion about picture perception. The issue of whether looking at pictures is similar as looking through windows has often risen (e.g., Gibson, 1971; Sedgwick, 1980; Wagemans et al., 2011b) but, until now, has never been experimentally addressed. Our data suggest that we should look for an alternative theory, which could be "the eye is not in pictorial space" that we named the (viewpoint independent) third hypothesis in the discussion of the first experiment. In the theory outlined by Koenderink and van Doorn (2008), the depth dimension of pictorial space is an affine line, that is, a Euclidean line without origin. The three dimensionality of pictorial space \mathbb{P}^3 is a product (officially called a "fiber bundle") of the Euclidean picture plane \mathbb{E}^2 and the affine depth dimensions \mathbb{A}^1 (see Koenderink & van Doorn (2012) for a recent theoretical account on this topic). The absence of a natural origin, or viewpoint, implies that objects with similar depths are located on a (affine) plane parallel to the picture plane. From this hypothesis, it follows that the equidistance loci in pictorial space are structured in *planes*, whereas in real space, they should be organized in curved surfaces (ideally spheres). Our results are certainly in this direction, but the theory cannot explain all of our results.

The difference between viewing pictures and reality appears too complex to be described by either of the two theories. Alberti's window hypothesis predicts no difference between pictures and reality, whereas the-eye-is-not-in-pictorial-space hypothesis only predicts categorical differences between pictures and reality, not within real (i.e., one eye vs. two eyes in Experiment 2) or pictorial (screens size effect in Experiment 1) conditions. Therefore, we cannot escape the conclusion that there appears to be a soft boundary between depth perception in pictures and in reality. Observers seem to use different strategies to infer depth when confronted with different ways of viewing. Pictorial space becomes more "real" when using a large display, and reality becomes more "pictorial" when we close one eye. Making reality pictorial has also been demonstrated to occur when we view a real scene through a frame. Eby and Braunstein (1995) framed a real scene that caused observers to experience a similar depth flattening as is usually found in pictorial space. The current goal of the display industry is to make pictorial space more like real space by increasing the sense of so-called presence (e.g., IJsselsteijn, de Ridder, Freeman, & Avons, 2000). Imaging systems are designed to convince the observer that he or she is present in the virtual environment. The feeling of presence is often investigated using questionnaires (Witmer & Singer, 1998) that probe a mental awareness, whereas our study takes the idea of presence literally: We measured whether observers are geometrically present. The results indicate that the amount of geometrical presence varies gradually between real and pictorial space. Large displays increase presence, but more extreme applications like head-mounted displays may eventually break the final border between real and pictorial space. Under normal circumstance and for conventional small screens, posters, and paintings, we can conclude that Alberti's window hypothesis does not hold. To speak with Marina Abramovic: "the artist is present," but the observer is absent.

References

- Alberti, L. B. (1970). On painting. New Haven, CT: Yale University Press. (Original work published 1435)
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357
- Cutting, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. Journal of Experimental Psychology: Human Perception and Performance, 13, 323–334. doi:10.1037/0096-1523.13.3.323
- Cutting, J. E. (2003). Looking into pictures: An interdisciplinary approach to pictorial space. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), *Reconceiving perceptual space* (pp. 215–238). Cambridge, MA: MIT Press.
- Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Handbook of perception and cognition. Volume 5: Perception of space and motion* (pp. 69–117). San Diego, CA: Academic Press.
- Da Vinci, L. (2004). *The notebooks of Leonardo Da Vinci Vol. 1*. Retrieved from http://www.gutenberg.org/ebooks/5000
- Eby, D. W., & Braunstein, M. L. (1995). The perceptual flattening of three-dimensional scenes enclosed by a frame. *Perception*, 24, 981–993.
- Gibson, J. J. (1954). A theory of pictorial perception. *Educational Technology Research and Development*, 2, 3–23.
- Gibson, J. J. (1971). The information available in pictures. *Leonardo, 4*, 27–35.
- Gombrich, E. H. (1960). Art and illusion: A study in the psychology of pictorial representation. London, UK: Phaidon Press.
- Haber, R. N. (1985). Toward a theory on the perceived spatial layout of scenes. Computer Vision, Graphics, and Image Processing, 31, 282– 321.
- Hecht, H., van Doorn, A., & Koenderink, J. J. (1999). Compression of visual space in natural scenes and in their photographic counterparts. *Perception and Psychophysics*, 61, 1269–1286.
- IJsselsteijn, W. A., de Ridder, H., Freeman, J., & Avons, S. E. (2000). Presence: Concept, determinants and measurement. *Proceedings of the SPIE*, 3959, 520–529.
- Kepes, G. (1995). Language of vision. New York, NY: Dover.
- Koenderink, J., & van Doorn, A. (2008). The structure of visual spaces. Journal of Mathematical Imaging and Vision, 31, 171–187.
- Koenderink, J. J., & van Doorn, A. J. (2012). Gauge fields in pictorial space. SIAM Journal on Imaging Sciences, 5, 1213–1233.
- Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1992). Surface perception in pictures. *Perception & Psychophysics*, 52, 487–496.
- Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2001). Ambiguity and the "mental eye" in pictorial relief. *Perception*, *30*, 431–448.
- Kosslyn, S. M., Pick, H. L., & Fariello, G. R. (1974). Cognitive maps in children and men. *Child Development*, 45, 707–716.
- Masheck, J. (1991). Alberti's "window": Art-historiographic notes on an antimodernist misprision. Art Journal, 50, 35–41.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Sedgwick, H. A. (1980). The geometry of spatial layout in pictorial representation. *The Perception of Pictures*, *1*, 33–90.
- Todd, J. T. (2004). The visual perception of 3D shape. Trends in Cognitive Sciences, 8, 115–121.
- Van Doorn, A. J., Koenderink, J. J., & Wagemans, J. (2011). Rank order scaling of pictorial depth. *i-Perception*, 2, 724–744.

Vishwanath, D., Girshick, A. R., & Banks, M. S. (2005). Why pictures look right when viewed from the wrong place. *Nature Neuroscience*, 8, 1401–1410.

Wagemans, J., Van Doom, A. J., & Koenderink, J. J. (2011a). Measuring 3D point configurations in pictorial space. *i-Perception*, 2, 77–111.

- Wagemans, J., Van Doorn, A. J., & Koenderink, J. J. (2011b). Pictorial depth probed through relative sizes. *i-Perception*, 2, 992–1013.
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin and Review*, 11, 192–196.

White, J. (1987). The birth and rebirth of pictorial space. Cambridge,

MA: Belknap Press.

- Wijntjes, M. W. A., & Pont, S. C. (2010). Pointing in pictorial space: Quantifying the perceived relative depth structure in mono and stereo images of natural scenes. *Transactions on Applied Perception*, 7, 1–8.
- Willats, J. (1997). Art and representation: New principles in the analysis of pictures. Princeton, NJ: Princeton University Press.
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7, 225–240.

Appendix

Reconstruction Algorithm

For *N* sample points z_i , we have N(N - 1)/2 pairs $\{z_i, z_j\}|_{i < j}$. For each pair, the observer indicates which point appears closer. We first assume that this depth difference is metric. From these relative depth differences, we want to calculate the overall depth, which can be formulated in a straightforward way,

$$Mz = \Delta z \tag{2}$$

which stands for

$$\begin{pmatrix} 1 & -1 & 0 & 0 & \cdots \\ 1 & 0 & -1 & 0 & \cdots \\ 1 & 0 & 0 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \\ 0 & 1 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ \vdots \end{pmatrix} = \begin{pmatrix} z_1 - z_2 \\ z_1 - z_3 \\ z_1 - z_4 \\ \vdots \\ z_2 - z_3 \\ z_2 - z_4 \\ \vdots \end{pmatrix}$$
(3)

The vector Δz contains the observers' responses. Equation 3 denotes a system of N linear equations of the form $A\mathbf{x} = \mathbf{b}$, which can be solved in a linear squares fashion by taking the pseudoinverse of Matrix A. When using the Moore-Penrose pseudoinverse method, we get the (perhaps surprisingly simple) solution:

$$z_i \simeq \frac{1}{N} \sum_{j \neq i}^N \left(z_i - z_j \right) \tag{4}$$

In prose, this equation means that the depth of a certain point is equal to the sum of depth differences with all other points. It is approximate because it is a least-squares solution. Intuitively, this equation makes sense if you imagine an infinite set of randomly distributed z_j . For each z_i , the large set of z_j will sum to the same fixed value for all z_i . This fixed value can be set to zero (it is an irrelevant depth offset), and one gets

$$z_i = \frac{1}{N} \sum_{j \neq i}^{N} (z_i) = \frac{z_i}{N} \sum_{j \neq i}^{N} 1 = z_i.$$

For our purpose, the difference $(z_i - z_j)$ is a binary response. Let us denote this response by r_{ij} , for which

$$r_{ij} = \begin{cases} 1 & if x_i > x_j \\ 0 & if x_i < x_i \end{cases}$$
(5)

Furthermore, we set $r_{ij} = 1 - r_{ji}$. Now Equation 4 can be written as

$$z_i \simeq \frac{1}{N} \sum_{i \neq i}^N r_{ij} \tag{6}$$

This equation means that the depth of a point equals the sum of the positive responses (or "votes") that that point is closer that another point.

> Received January 24, 2013 Revision received September 3, 2013

Accepted September 9, 2013