Depth perception in 3D clutter

Michael Langer*, Haomin Zheng, Shayan Rezvankhah

School of Computer Science

McConnell Eng. #318, 3480 University

McGill University

Montreal Canada

*Correspondence should be sent to michael.langer@mcgill.ca

KEYWORDS: depth perception, binocular stereopsis, motion parallax, natural scenes

Abstract

Many natural objects such as trees, shrubs, and tall grass consist of thousands of small surfaces that are distributed over a 3D volume. A common vision task in 3D cluttered scenes is to estimate the depth of objects within such 3D clutter. For example, one may need to decide whether a fruit is reachable, or what is the distance to a predator or prey. To perceive depth in 3D clutter, an observer combines several cues. Two of the most important depth perception cues are binocular disparity and motion parallax. These two cues are less reliable in 3D clutter, however, since surfaces may be partly occluded. Occlusions themselves contain depth information, but it is unknown if visual systems use this information in 3D clutter, as previous studies have considered occlusions for simple scene geometries only. Here we present a human vision experiment that addresses this question. We identify two new metric depth cues that are available in 3D clutter. The first is based on the fraction of an object that is visible, and the second is based on the depth of the occluders. Our experiment shows that the visual system uses both of these cues, along with binocular disparity and motion parallax. The net effect is that depth perception suffers very little from occlusions in 3D clutter. The 3D clutter itself provides depth information that compensates for a loss of information from binocular stereo and motion parallax.

Introduction

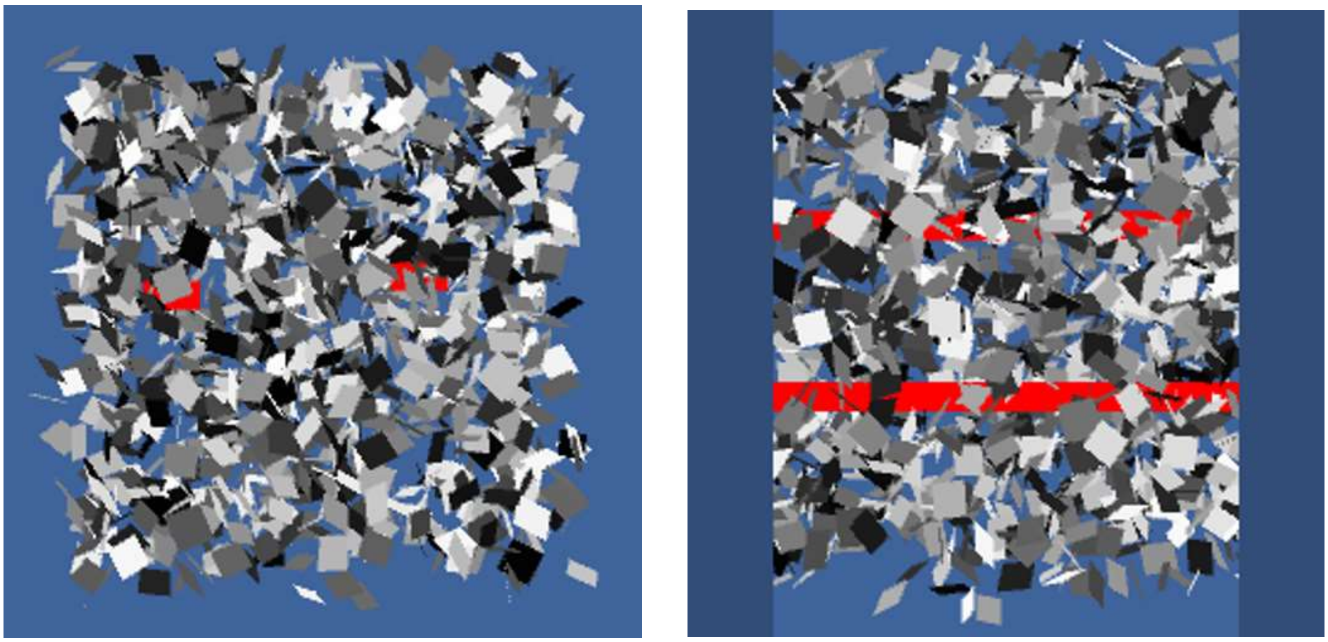
The human visual system has evolved over millions of years, predominantly in cluttered 3D environments such as forest and grassland. Such environments consist of objects such as trees, shrub, and tall grass that contain thousands of individual surfaces scattered in 3D space. Such 3D clutter leads to objects being partly occluded which reduces visibility [Changizi and Shimojo, 2008] and complicates the task of depth perception. In particular, visual cues such as binocular disparity and motion parallax which normally play an important role in depth perception are less reliable in 3D clutter since the occlusions make it more difficult to find corresponding points, between the eyes in the case of binocular stereopsis or over time in the case of motion parallax.

Previous studies of depth perception in 3D clutter have concentrated on the depth of the clutter elements themselves. Two types of questions have been addressed. The first question is how many discrete depth planes can the visual system perceive, for example, from cues such as binocular disparity [Akerstrom and Todd 1988, Tsirlin 2008] or motion parallax [Andersen 1989]. The second question is how well can the visual system judge the depth-to-width ratio of the 3D clutter [van Ee and Anderson 2001, Harris 2014]. Both types of study have provided insight into the limitations of binocular disparity and motion parallax cues in 3D cluttered scenes. However, these studies are incomplete since they ignore an essential aspects of depth perception in 3D clutter, namely these studies ignore occlusions. They do so by assuming the clutter consists of sparse elements only, typically white points or lines seen against a black background, or vice-versa, black points seen against a white background.

Our goal was to address a different 3D clutter scenario in which the objects are 2D surfaces that produce dense occlusions. This is often the case in foliage, for example. We explicitly examine the information provided by occlusions. Moreover, rather than examining the depth distribution of the clutter directly, we examine how well observers can perceive the depth of objects that are located within the 3D clutter.

Consider the 3D cluttered scenes shown in Figure 1a and 1b. These scenes are an abstraction of natural foliage consisting of large number of leaves and a few identifiable objects such as fruit or horizontal tree branches. Specifically the clutter consists of gray distractors that are distributed over a volume, and two red targets that have different depths within the volume. In Fig. 1a, the two targets are short red bars that are positioned in the left and right halves of the volume. In Fig. 1b, the targets are long red bars that are positioned in the upper and lower halves of the volume. For each scene, the observer's task is to discriminate the depth of the targets, namely to decide which of the two targets is shallower in depth, that is, closer to the observer. The scenes are presented using an Oculus Rift head mounted display which allows us to control whether the binocular disparity and motion parallax cues are present (see Methods).

The goal of the experiment was to study the information that is available beyond the traditionally studied binocular disparity and motion parallax cues. We have identified two new novel metric depth cues that are present in 3D clutter. The first is a 'visibility' cue which is based on a probabilistic relationship between the depth of a target and the visibility of that target [Langer and Mannan 2012] namely, for a given 3D clutter density, the target that is more visible is more likely to be closer to the observer. The second cue is based on the depths of the occluders which can be perceived using binocular disparity and motion parallax cues. We refer to this as the occluder depth cue or, more simply, the 'occlusions' cue. The idea is that when a target is partly occluded and when the depth of the occluder can be perceived, this occluder depth provides a lower bound on the depth of that target. When there are several occluders for a target, the target depth must be greater than the maximum depth of these occluders. This depth bound provides metric depth information, and it is distinct from the figure-ground metric depth from occlusions cue which has been identified previously [Burge et al, 2010].



(a) (b)

Figure 1. Each rendered scene consisted of two red targets that are embedded in a 3D field of distractors. The targets were rectangular (a) short bars separated horizontally and in depth or (b) long bars separated vertically and in depth. The target surfaces always faced the Z direction and the height of each target was always 1 degree of visual angle. The short bar targets had a horizontal: vertical aspect ratio of 2:1. The long bar targets extended beyond the width of the clutter. The left and right edges of the long bar targets were hidden behind large flanking vertical occluders which removed binocular disparity and motion parallax cues.

Methods

Apparatus

We used an Oculus Rift DK2 binocular head mounted display. Scenes were rendered using Unity, a cross-platform game engine developed by Unity Technologies, and using C# as the scripting language. The software ran on a Dell Precision T7610 equipped with an NVIDIA Quadro 4000K graphics card. Scenes were rendered in real-time using a head coupled perspective model of the observer's left and right 3D eye positions [Sutherland 1968, Arthur et al 1993]. The Rift comes with a motion sensing camera and accelerometer for position tracking, and a gyroscope and magnetometer for orientation tracking. For a description of the Oculus Rift DK1 version, see [Lavalle 2014]. Observer position tracking was achieved using the Unity plugin provided in Oculus Rift SDK. The position update rate for the Rift is 60 Hz and the display refresh rate is 75 Hz. The Rift has an OLED display with a resolution of 1920 x 1080 pixels (960 x 1080 per eye), and a horizontal field of view of 100 degrees, or about 10 pixels per degree. This is a much lower resolution for a desktop monitor, but it was sufficient for subjects to do the task.

Stimuli

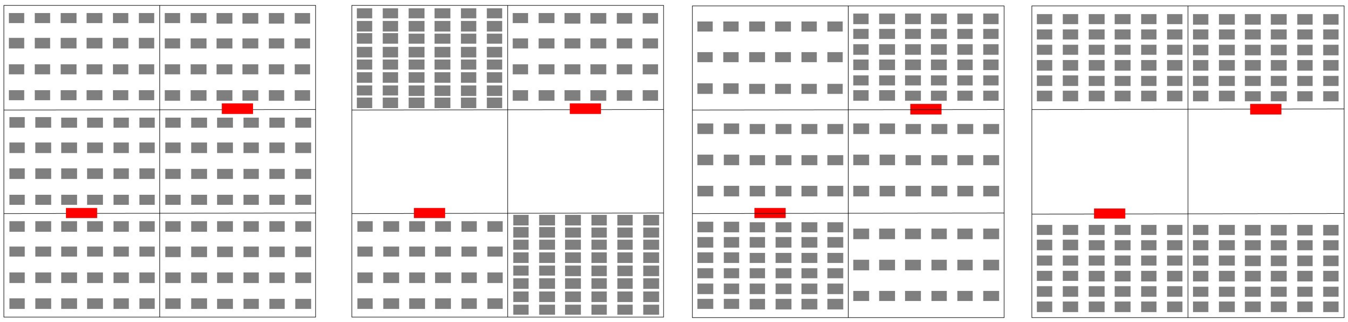
On each trial, the XYZ positions of the two targets were chosen as follows. The centers of the targets were initially placed at the center of a bounding XYZ volume of size 20 x 20 x 20 cm which was centered at the origin. The short bar targets were separated horizontally (X) by 10 cm, and then the XY position of each target was perturbed slightly to reduce positional cues to depth. The long bar targets were separated vertically by 6.7 cm. Both short and long bars targets then were separated in depth by an interval ΔZ which was chosen using a staircase procedure that will be described below. This depth interval always was centered at the middle depth of the volume.

In each trial, there were 1331 (11 x 11 x 11) new distractors defined. Each distractor was initialized to be a square of width 12 mm, and assigned a random grey level reflectance. (Only ambient lighting was used, that is, no shading.) The 3D orientation of each distractor was chosen randomly. The distractor then was placed within the XYZ bounding volume according to one of the distributions that we define in the caption of Figure 2. The front face of this XYZ bounding volume was at a distance of 60 cm from the viewer and thus the clutter lay in the depth interval [Zmin,Zmax] = [60,80].

One subtle detail should be mentioned about the distractor positions. For each distractor, a candidate XYZ position was generated and it was then determined if that distractor intersected one of the two targets. If it did, then the distractor depth was shifted slightly in depth to avoid this intersection. The reason for avoiding intersections between distractors and targets is that the intersection points would provide features that could be used for depth from binocular disparity or motion parallax. It was important to remove these features in order to study the two new depth cues (visibility and occlusions) in isolation. In particular, the purpose of using the long bar conditions was to have no direct information about depth of the targets from binocular disparities and motion parallax.

Scenes were rendered under perspective projection. The visual angle of each distractor naturally varied with inverse depth which provides a size cue for the distractors. We removed the size cue for the targets by scaling each target so that its projected image size was the same in all trials, namely the height of each target corresponded to that of a 12 mm high target at 70 cm depth. The reason we removed the size cue for the target was so that we did not to confound it with the visibility cue. See Discussion later.

Each scene was displayed either stereoscopically or monoscopically, and with or without head coupled perspective. For simplicity, we refer to these conditions as “stereo”or “mono” and “motion parallax” or “no motion parallax”, respectively. For the stereo + motion parallax condition, the scene was rendered separately for each eye, using the eye’s actual 3D position which was estimated in real time by the Rift. An interocular distance of 6.4 cm was used for all observers. For mono viewing, the scene was rendered from the midpoint between the eyes and the same image frames were shown to both eyes. For the motion parallax conditions, observers were instructed to move their heads left and right. We clipped the rendered observer’s position to a horizontal XYZ line segment of size 30 x 0 x 10 cm which was centered at position (0, 0, 60) relative to the center of the front face of the clutter cube. This restricted the viewing position to always have the same y value, which removed any possibility that the observers could use vertical motion parallax from the target's long horizontal edges. In the no motion parallax conditions, the scene was rendered from the standard viewing position, namely (0, 0, 60).



(a) Visibility + Occlusions (b) Visibility only (c) Occlusions only (d) None

Figure 2. Four combinations of visibility and occlusion cues. These combinations are illustrated for an XZ slice through a short bar scene or YZ slice through a long bar scene. For each panel, the near target has depth Znear = 70 - $\Delta Z/2$ cm and the far target has depth Zfar = 70 + $\Delta Z/2$ cm. The distractors are drawn in uniform grey and regularly spaced to illustrate how the density properties vary. In the experiment, the grey level, position, and orientation were randomized. **(a) Visibility + Occlusions:** Each distractor's Z position is chosen according to a uniform distribution over the Z range of the volume. **(b) Visibility only:** The distribution is similar to (a) except that, for the half volume containing the near target, we move the distractors from the depth interval [Znear, Zfar] to the depth interval [Zfar, Zmax] and, for the half containing the far target, we move the distractors from the depth interval [Znear, Zfar] to the depth interval [Zmin, Znear]. Like in (a), there is a visibility cue since the expected number of occluders for the near target is less than the expected number for the far target. Unlike in (a), there is no occlusions cue since the depth range of foreground occluders is the same for the near and far targets. **(c) Occlusions only:** For each half volume, the probabilities that a distractor has depth less than the target or greater than the target are each 0.5. There is no visibility cue because the expected number of distractors with depth less than the target is the same in both halves and so the expected visibilities of the two targets are the same. There is an occlusions cue since the distractors that can occlude the near target have a smaller depth range than the distractors that can occlude the far target. **(d) None:** Each distractor appears at a depth less than Znear with probability 0.5 and at a depth greater than Zfar with probability 0.5. There is no visibility cue because the expected visibility of the two targets is the same. There is no occlusion cue because the occluders only constrain both targets to lie in the same interval [Znear, Zfar].

Design

The four cues - binocular disparity, motion parallax, visibility, occlusions – could each be turned on or off, which defined 16 possible conditions. We removed two of these conditions for the short bar targets and we removed five of these conditions for the long bar targets. These were conditions in which the task was impossible. For the short bar targets, there were two such conditions namely when there were no stereo, motion, or visibility cues. The occlusions cue could either be present or not in this case. That is, the task was impossible even if there was an occlusion cue, since this cue requires that the depth of the distractors can be perceived which requires at least one of the stereo or motion parallax cues. Strictly speaking, there was a size cue for the distractors as well as ordinal occlusion cues and visibility cues for the distractors which provided some depth information. However, in practice these cues are very weak for the distractors and in pilot studies we found that the performance was at chance when only the occlusions cue was available for the target. For the long bar targets, pilot studies showed that the task was also impossible when there were no stereo, motion, or visibility cues for the same reason as for the short bar targets. In addition, the task was impossible for the long bar targets when there were stereo and/or motion cues, but there were neither visibility nor occlusion cues. The reason is that the stereo and motion cues provide depth information only about the distractors in this case. For these two reasons, we removed another three conditions for the long bar targets. Note that the condition in which all four cues were absent had already been removed.

We also tested a baseline condition for the short bar targets in which stereo and motion cues were present, but all distractors were removed. This gave a total of 26 conditions, namely 25 clutter conditions and one baseline condition. We did not include a baseline condition for the long bar targets since there was no information about depth for these targets, i.e. when the two large flanking occluders were present.

Each observer ran all 26 conditions in a blocked design, with one staircase per block. The ordering of the blocks was randomized for each observer.

Observers

Fifteen observers participated. Each was a student at McGill University and was paid $10. Observers had little or no experience with psychophysics experiments. Each had normal or corrected-to-normal vision. We required that each observer could discriminate 50 arcsec of disparity to participate, namely level 6 of the Randot Stereo Test (Precision Vision). Observers were unaware of the purpose of the experiments. Informed consent was obtained using the guidelines of the McGill Research Ethics Board which is consistent with the Declaration of Helsinki.

Procedure

In each trial, the task was to discriminate the depths of the targets, namely to indicate which of the two targets was closer to the observer. Observers responded by the pressing keys on the keyboard: left-right arrows for short bar targets, and up-down arrows for the long bar targets.

A blocked design was used with one condition for each block. For each block, a one-up/one-down staircase was used with different step sizes for down steps versus up steps. The ratio between the log of the up-step size and the log of the down-step size was chosen as 0.2845 [Garcia-Perez, 1998]. This ratio aims for approximately 78 percent correct. Specifically, whenever the subject answered correctly, we reduced the distance ΔZ between targets by a factor 0.8, and when the observer answered incorrectly we increased ΔZ by a factor 2.19. Each staircase began at level $\Delta Z = 12$ cm and terminated after 12 reversals. To compute the threshold for a given staircase, we averaged the log of the ΔZ values for the last 10 reversals. If ΔZ increased beyond 20 cm which normally would put the targets outside the bounding box of the clutter, we instead displayed the near target just in front of the front face at Zmin and the far target just beyond the back face at Zmax. This made the task trivial since the near target was unoccluded and the far target was highly occluded. If the observer still answered incorrectly in this case, we used the usual rule for choosing the next staircase level but again displayed the targets at the same depths Zmin and Zmax.

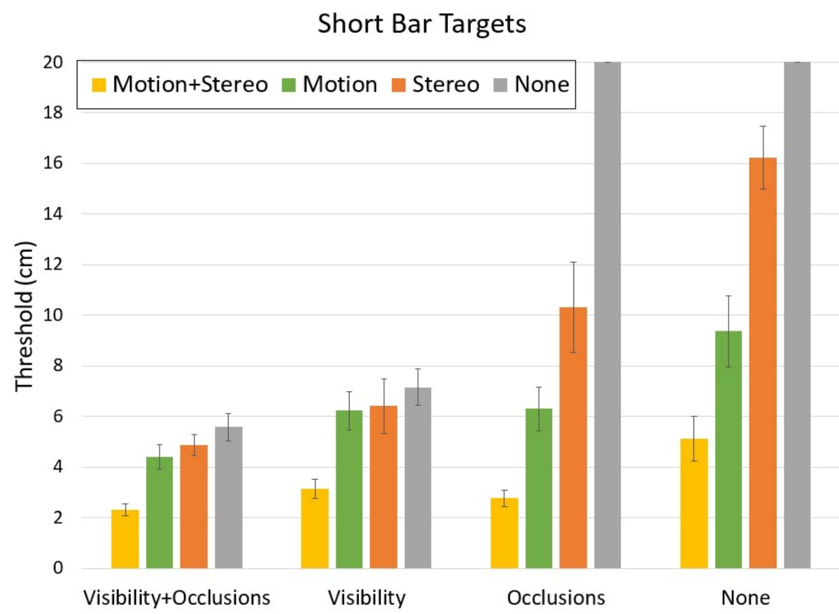
The response time in each trial was limited to four seconds. If the subject didn't respond, another scene was generated using the same target distance and a prompt was displayed to remind the subject to respond in time. For blocks in which there was no motion parallax cue, a warning message was presented telling subjects moved their heads. The message reminded subjects not to move, and the offending trial was discarded. Similarly, for blocks in which there was a motion parallax cue, a warning message was presented if subjects did not move their heads. The experiment typically lasted close to one hour.

Before running the experiment, each subject ran a very short practice session with three conditions, each with stereo present: the short bar targets with and without motion parallax, and the long bar targets with motion parallax. There was no time limit in each trial of the practice session. As in the real experiment, the initial ΔZ was 12 cm and a staircase was used to determine the next level. Since the purpose of the practice session was merely to familiarize the subjects with the requirements of the task, we kept the session short: each condition terminated with the first incorrect answer.

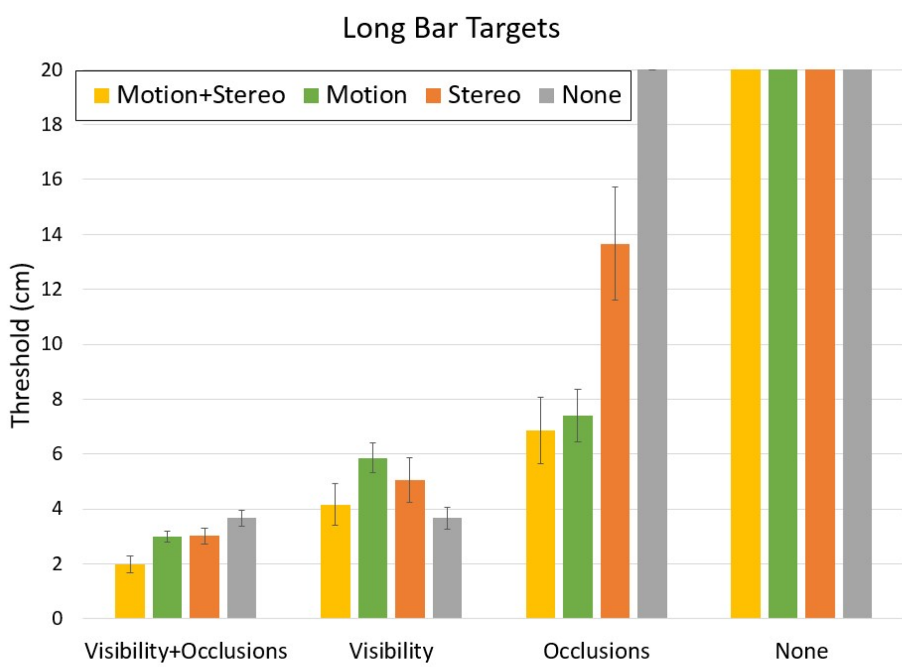
Results

To control the visibility and occlusion cues and to study how these cues interacted with binocular disparity and motion parallax cues, we manipulated the 3D distributions of the distractors (see Figure 2 caption). We measured depth discrimination thresholds for 15 naïve subjects, for various combinations of these four depth cues.

Figure 3a shows the results for the short bar targets. For each combination of binocular disparity and motion parallax cues, thresholds were lower when either the visibility or occlusion cues were present than when neither was present. These results show clearly that observers used the visibility and occlusion cues, even when depth information was available from binocular disparity and/or motion parallax. Thresholds were lower when both visibility and occlusion cues were present than when just one was present although the benefit from having both cues was not always significant. Unsurprisingly, we also found that thresholds were lower when both binocular disparity and motion parallax cues were present than when just one of these cues was present which is consistent with previous studies [Sollenberger 1993, Arthur et al 1993, Johnston et al 1994, Bradshaw and Rogers 1996]. Finally, when all four cues were present, the threshold was only slightly greater than the threshold in a baseline condition in which there were no distractors, although the difference was not statistically significant. Thus, although 3D clutter reduces the reliability of binocular disparity and motion parallax cues in this depth discrimination task, the reduction was almost fully compensated for by the visibility and occlusion cues.



(a)



(b)

Figure 3. Results for the (a) short bar targets and (b) long bar targets. A threshold of 20 cm is shown for conditions in which the task is impossible ΔZ is less than the depth interval of the clutter, and trivial when ΔZ is greater than the depth interval of the clutter. Error bars show standard error of the mean. "Motion" refers to conditions with the motion parallax. "Stereo" refers to conditions with binocular disparity.

The long bar target condition in Figure 1b was designed to remove all binocular disparity and motion parallax information about the targets themselves, by hiding the left and right target edges behind large flanking occluders. The only depth information about the long bar targets comes from the visibility and occlusion cues. Results are shown in Figure 3b. The general trend is similar to the short bar case, namely for each combination of binocular disparity and motion parallax cues, thresholds were lower when either the visibility or occlusion cue was present than when neither was present, and thresholds were lower when both visibility and occlusion cues were present than when just one was present.

As in the short bar case, for most combinations of visibility and occlusion cues, thresholds were lower when either binocular disparity or motion parallax cues were present. An interesting exception occurred in the 'visibility + no occlusions' condition, where thresholds were lowest when neither binocular disparity nor motion cues were present. This result is surprising at first glance since one normally expects binocular disparity and motion parallax cues to improve performance. The result can be understood by noting that, when binocular disparity or motion cues are present and when ΔZ is large, observers can perceive that the 3D density of the foreground distractors is different between the two targets. Since the visibility cue is only valid for the depth discrimination task when the 3D distractor densities for the two targets are roughly the same, it would make sense that the visual system relies less on the visibility cue when the distractor densities are perceived as being more different. Simply put, observers are more likely to guess in this condition which leads to an increased threshold. When neither binocular disparity nor motion parallax cues are present, observers have no way of knowing that the densities are different between the two targets, and in this case they rely heavily on the visibility cue and perform quite well.

Ideal Observers

To better understand the information available from visibility and occlusion cues, we defined ideal observers for these cues. These ideal observers are expected to perform better than the human observers since they use all the information available from these cues.

The ideal observers were defined as follows. In each trial, each ideal observer cast a set of rays to a regular grid of positions on the two targets, namely one ray was cast for each 0.5 x 0.5 mm^2 area on each target. The visibility of each target was defined as the fraction of these cast rays that did not intersect a distractor. The visibility-based ideal observer chose the target with larger visibility to be closer. The binocular version of this ideal observer summed the visibilities of the two eyes. The occlusions-based ideal observer was defined using the same set of cast rays. For each target, it considered the complementary set of rays, namely the cast rays that hit an occluder. This observer then computed the maximum Z value of those occluder rays, so this maximum Z value was a lower bound on the target depth. It then selected the closer of the two targets to be the one with smaller lower bound on target depth. The binocular version of the occlusions-based ideal observer pooled the rays for the left and right eyes and computed the lower bound based on the pooled occluder rays.

The method of constant stimuli was used. For each of the four conditions, namely {visibility only, occlusions only} x {mono, stereo}, percent correct scores were computed for twenty levels of ΔZ and 5000 trials per level. Four psychometric curves were obtained for the short bar targets and for the long bar targets. A 75% threshold was obtained by inspecting each of the eight psychometric curves.

We first consider the mono ideal observers. The visibility-based mono ideal observer performed worse for the short bar targets than for the long bar targets which was expected since there are fewer independent samples for computing visibility for each short bar target than for each long bar target, and so for any depth the computed visibilities of the short bar targets have a greater variance. Specifically, thresholds for the visibility-based ideal observers were 5 cm and 2.1 cm for the short and long bar targets, respectively. Thresholds for the human observers in the corresponding condition (visibility + no occlusion + no stereo + no motion) were only slightly greater, namely 6.5 cm and 3.8 cm respectively. The small difference between ideal and human observers in each case suggests that humans made near full use of the visibility cues.

The occlusion-based mono ideal observers performed better than the visibility-based ideal observers. Thresholds for the occlusion-based ideal observers were 0.6 cm and 0.2 cm for the short and long bar targets, respectively. This performance level is more difficult to relate to human performance levels, however, since the occlusion-based ideal observer has much more information for the doing the task than the human observers, namely the occlusion-based ideal observer is given the depths of each target's occluders whereas human observers must estimate the depths of the occluders from binocular disparity and/or motion parallax cues when they are present.

Interestingly, the binocular versions of the ideal observers performed almost identically to the monocular ones. Thresholds were lower by less than 0.1 cm for both types of binocular ideal observers and for both short and long bar target conditions. To understand why having a second eye gives so little benefit the two types of ideal observers, we make two observations. First, the left and right views of any target are typically similar since the only differences between the views is due to small disparity shifts of the occluders. Second, the near versus far target in a trial typically appear quite different, since the two targets have different depths and a different set of occluders. The net result of these two observations is that the left and right eye ideal observers typically would give the same response on each trial, correct or incorrect. This would be the same response as the binocular ideal observer who combines the left and right views and it also would be the same response as a monocular ideal observer positioned between the left and right eye. Since the binocular and monocular observers tend to agree on each trial, we expect them to have similar overall performance which is what we found.

We did not define ideal observers in the motion parallax case, but presumably the performance would be similar to the binocular case, assuming the moving viewer's position was restricted to the line segment joining the two eyes. Motion parallax should provide slightly more information that binocular disparity since it provides a continuum of views rather than two discrete views. However the extra information from multiple views will be limited in the same way, namely that the differences between images of one target across multiple views typically would be much less than the differences between the images of the two targets from one view. To summarize, very little extra information about depth from visibility or occlusions is provided by having multiple views over having a single view, when the viewpoints are relatively near each other in space.

Discussion

We have shown that humans use the visibility and occlusion cues along with binocular disparity and motion parallax cues to discriminate depth in 3D clutter. We have concentrated mostly on the visibility and occlusion cues, and examined some of the interactions between these cues and binocular disparity and motion parallax. Here we turn to our attention to the latter two cues and consider the detailed information that is provided by each of these cues in 3D clutter. We also discuss other cues that could contribute to depth perception in 3D clutter.

The binocular disparity cue takes at least three distinct forms in 3D clutter, all of which are present for the short bar targets but not the long bar targets. The first is the classical disparity cue, namely position differences of corresponding points in the left and right eye views. In our experiment, the only corresponding points on the short bar targets are on the side edges. The second and third forms of binocular disparity in 3D clutter are more speculative. The second form is da Vinci stereopsis [Nakayama and Shimojo 1990, Harris and Wilcox 2009] which occurs when a side edge of a target is visible to one eye but not to the other. Previous studies of da Vinci stereopsis have considered only simple scene geometries, but the cue is available in principle in 3D clutter as well. Note that da Vinci stereopsis only informs the observer about the depth of the target relative to its occluder, however, so it would be useful for our depth discrimination task only if the occluder's depth can be perceived reliably, which is less likely in 3D clutter. The third form of the disparity cue is the binocular position differences of those target points for which a unique left/right correspondence cannot be found. For the short bar targets, these would be points other than those on the side edges. The visual system could estimate the disparity by comparing the envelopes of these points in the left and right images, although to our knowledge this has never been studied before. This envelope cue is reminiscent of the disparity of Gaussian target envelopes that are used in studies of second order stereopsis [Wilcox and Allison, 2009]. Perhaps these second order mechanisms are used in cluttered scenes as well.

It is well known that motion parallax cues are similar to binocular disparity cues [Rogers and Graham, 1982] and so we can list three forms of motion parallax cues as well. The first is classical motion parallax, namely when an observer moves laterally, points on the side edges of short bar targets move relative to each other, which provides information about their relative depth. The second form of motion parallax is a motion analogue of da Vinci stereopsis. Suppose a target were partly occluded and the occluder were visible over the duration of a head movement such that the occluder's depth in the volume could be perceived reasonably accurately. Moreover, suppose that the target's side edge became visible for a fraction of the duration of the observer's head movement, which is not long enough for the target's velocity to be perceived reliably. In this case, the relative depth between the occluder and target might still be perceivable using the width of the visible region of the target during the brief time that the target is visible. Whether or not the visual system uses this motion analogue of da Vinci stereopsis has not been investigated before to our knowledge. The third form is the motion of target points whose image velocity is not uniquely defined in the stimuli, namely points other than those on the side edges of the short bar targets. Observers could perform the depth discrimination task by comparing the velocities of the envelopes of the two targets. This envelope motion is reminiscent of stimuli used to study second order motion in human vision [Chubb and Sperling 1988], and it is plausible that observers use second order motion mechanisms to make parallax judgements in clutter as well.

We next turn to depth cues that did not play an explicit role in our experiment, but that should be considered in more general studies of depth perception in 3D clutter. The first is the classic 'size cue', namely the image size of a target. Image size varies directly with depth according to linear perspective and, in the absence of occlusions, image size is identical to the visible solid angle since all points on the target are visible. In 3D clutter, however, the expected visible (solid) angle of the target varies directly with depth, and thus the image size and expected visible solid angle are statistically correlated. We removed the size cue from our stimuli by scaling each 3D target based on its depth so that the image size would be constant in the absence of occlusions. We did so to ensure that observers were using the fraction of the target that was visible, rather than the size cue. However, in a more general situation in which the size cues is also present, the visibility and size cues will be confounded and observers will use both.

Another cue that is often present in 3D clutter is shading and shadows. Surfaces that are deeper within a 3D volume tend to receive less illumination and hence appear darker since the volume itself occludes the light coming from the scene. This 'dark means deep' cue been shown to play a role in perception of shading on smooth surfaces [Langer and Buelthoff 2000]. This cue also is the basis of the 'ambient occlusion' method which is commonly used in volume rendering in computer graphics [Diaz et al 2015]. It would be interesting to examine how how the visual system weighs such shading cues with the other depth cues we have examined.

In conclusion, our experiment has provided new and fundamental insights into depth perception in 3D clutter, in particular, in situations where the clutter is dense and occlusions effects are significant. We have identified two new metric cues to depth in 3D clutter: a visibility cue and an occluder depth cue. We have shown how humans combine these depth cues with binocular disparity and motion parallax. Although one might have expected that 3D clutter simply interferes with depth perception by reducing the information from binocular disparity and motion cues, we have shown this is not the case. Rather, 3D clutter also can aid depth perception of targets by providing these two new metric depth cues, even in situations such as our long bar targets where binocular disparity and motion parallax cues provide no direct depth information about targets. Our 3D clutter scenario opens a rich, new, and natural domain for studying depth cue combinations, using the four cues we examined as well as other cues such as size and shading.

Acknowledgements

This research was supported by a grant from FQRNT (Fonds québécois de la recherche sur la nature et les technologies).

References

Akerstrom, R. A., Todd, J. T. (1988). “The perception of stereoscopic transparency.” Perception and Psychophysics 44, 5, 421–432.

Andersen, G. 1989. “Perception of three-dimensional structure from optic flow without locally smooth velocity.” Journal of Experimental Psychology: Human Perception and Performance 15, 2, 363–371.

Arthur, K. W., Booth, K. S., Ware, C. 1993. “Evaluating 3d task performance for fish tank virtual worlds.” ACM Transactions on Information Systems 11, 3, 239–265.

Bradshaw M. F., Rogers B.J. 1996, “The Interaction of Binocular Disparity and Motion Parallax in the Computation of Depth”, Vision Research, 36(21): 3457-3468

Burge, J., Fowlkes, C. C., Banks, M. S. (2010). “Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception.” The Journal of Neuroscience, 30, 7269–7280

Changizi MA and Shimojo S. (2008), “X-ray vision and the evolution of forward-facing eyes,” Journal of Theoretical Biology 254, 756–767.

Chubb, C, Sperling, G (1988). "Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception". J. Opt. Soc. Am. A 5 (11): 1986–2007

Díaz J, Ropinski T, Navazo I, Gobbetti E, Vázquez P. (2015) “An experimental study on the effects of shading in 3D perception of volumetric models”. The Visual Computer 31: 1-15.

Garcia-Perez, M. A. (1998). “Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties.” Vision Research 38 (12), 1861–1881.

Harris J., Wilcox L. (2009). “The role of monocularly visible regions in the perception of three-dimensional scenes,” Vision Research 49, 2666–2685.

Harris, J. M. (2014). “Volume perception: Disparity extraction and depth representation in complex three-dimensional environments.” Journal of Vision 14, 12, 1–16.

Johnston E.B., Cumming B.G., Landy M.S. (1994), “Integration of stereopsis and motion shape cues”, Vision Research, 34, pp. 2259–2275

Langer M.S., Buelthoff H. H (2000), “Depth discrimination from shading under diffuse lighting”, Perception.*29 (6) 649-660.*

Langer M. S., Mannan, F. (2012). “Visibility in three dimensional cluttered scenes.” Journal of the Optical Society of America A 29, 9, 1794–1807.

LaValle SM, Yershova A, Katsev M, Antonov M (2014), "Head tracking for the Oculus Rift", *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 187-194.

Nakayama K., Shimojo, S. (1990). "Da vinci stereopsis: Depth and subjective occluding contours from unpaired image points". Vision Research 30, 11, 1811–1825.

Rogers B, Graham M. (1982) "Similarities between motion parallax and stereopsis in human depth perception." Vision Research 22(2):261-70.

Sollenberger, R. L., Milgram, P. (1993). "Effects of stereoscopic and rotational displays in a three-dimensional pathtracing task." Human Factors: The Journal of the Human Factors and Ergonomics Society 35, 3, 483–499.

Sutherland, I. E. (1968). "A head-mounted three dimensional display." In Proceedings of the AFIPS Fall Joint Computer Conference, 757–764.

Tsirlin, I., Allison R. S., and Wilcox, L. M. (2008). "Stereoscopic transparency: Constraints on the perception of multiple surfaces." Journal of Vision 8, 5, 1–10.

van Ee, R., Anderson B., (2001) "Motion direction, speed, and orientation in binocular matching," Nature 410, 690–694.

Wilcox L. M., Allison R. S., (2009) "Coarse-fine dichotomies in human stereopsis", *Vision Research*, 49, pp. 2653-2665.