PSIHOLOGIJA, 2010, Vol. 43 (4), (389–409)

UDC 159.937.52.072 ; 159.938.3
DOI: 10.2298/PSI1004389V

Influence of depth cues on multiple objects tracking in 3D scene

Vesna Vidaković[1] and Sunčica Zdravković[1, 2]

[1]*Department of Psychology, Faculty of Philosophy, University of Novi Sad, Serbia*
[2]*Laboratory of Experimental Psychology, Faculty of Philosophy, University of Belgrade, Serbia*

Multiple-object-tracking tasks require an observer to track a group of identical objects moving in 2D space. The current study was conducted in an attempt to examine object tracking in 3D space. We were interested in testing influence of classical depth cues (texture gradients, relative size and contrast) on tracking. In Experiment 1 we varied the presence of these depth cues while subjects were tracking four (out of eight) identical, moving objects. Texture gradient, a cue related to scene layout, did not influence object tracking. Experiment 2 was designed to clarify the differences between contrast and relative size effects. Results revealed that contrast was a more effective cue for multiple object tracking in 3D scenes.

The effect of occlusion was also examined. Several occluders, presented in the scene, were occasionally masking the targets. Tracking was more successful when occluders were arranged in different depth planes, mimicking more natural conditions. Increasing the number of occlusions led to poorer performance.

Key words: *multiple object tracking, texture gradients, relative size, contrast, depth cues*

The human visual system, in addition to being able to detect and recognize a large number of objects, is also capable of simultaneously tracking several moving targets. In our everyday experience, these target objects could be located in very different parts of the visual field, move at different speeds and/or in different directions. They could also change depth in three-dimensional space, approaching or receding, and may often become partially occluded. Our ability to function in complex, dynamic environments – for example when crossing a city street or playing a competitive team sport — suggests that we can achieve such tracking quite effectively and with little effort. Since it is impossible to simultaneously examine all factors that might affect tracking in everyday

Corresponding author: szdravko@f.bg.ac.rs

This research was supported by the Ministry of Science and Technological Development of Serbia, grant number D–149039.

conditions, particularly as many of the factors may be closely linked, researchers have developed an experimental task that can be used to systematical control parameters of interest. This basic method is known as the Multiple Object Tracking (MOT) paradigm (Pylyshyn & Storm, 1988).

Multiple object tracking – MOT

Pylyshyn and Storm (1988) developed the MOT paradigm to examine the tracking of randomly moving objects. They showed that participants could easily track five moving targets presented together with another five identical objects and to accurately indicate them after the motion had ceased. Pylyshyn and Storm (1988) proposed that every moving object in the visual field was connected to one reference token. This connection made the tracking task possible. The reference tokens were compared to the fingers we use to point at objects and so were called FINSTs (FINgers of INSTantiation). According to this model, the cognitive system has limited numbers of FINSTs and consequently only a limited number of moving objects (3–5) could be processed simultaneously. Based on these and other findings, researchers have estimated four as the optimal number of mobile targets in most tracking tasks (i.e. Scholl, Pylyshyn, & Feldman, 2001; Oksama & Hyönä, 2004).

Although the basic experimental paradigm has been widely accepted and used, the original theoretical explanation put forward by Pylyshyn & Storm (1988) has been challenged. For example, an alternative approach was proposed by Yantis (1992) who suggested that separate moving elements were grouped into one virtual object, which was tracked as a unique form. The success of tracking should thus depend on factors influencing grouping into a unique form, as well as maintenance of the formed group.

In the present study we also exploit the MOT paradigm, but were particularly interested in more complex viewing conditions. That is, in order to reveal some of the strategies the visual system might utilize when faced with tracking in the real world, we felt that we had to provide more natural viewing conditions. Below, we briefly review a range of other studies that have included more complex stimuli in the context of MOT, before describing our own displays in detail.

MOT – a brief review

In the basic experimental paradigm, designed by Pylyshyn and Storm (1988), all ten of the tracked objects were identical. This object was a white „plus“ symbol (+) on a dark computer screen and up to five objects were marked as targets. All objects were initially stationary, and the target objects were marked during the introductory phase. This marking was removed and all objects started to randomly move across the screen. Occasionally one object would be highlighted and participants were asked to respond if this object was one of the originally selected targets.

This basic paradigm has been gradually changed in an attempt to bring the task closer to natural conditions. For example, moving objects found in natural scenes are seldom identical or equidistant, with well-separated trajectories. Also, in everyday viewing conditions, moving objects are rarely continuously visible. Instead they are often occluded with different types of the obstacles (Kellman & Shipley, 1991). We begin our review by looking at the role of occulders, before we include the MOT paradigm into more complex 3D context.

MOT with occluders

To investigate tracking and occlusion, Scholl and Pylyshyn (1999) conducted a series of experiments in which moving objects would disappear behind the occluders. Their results revealed that: (1) object tracking is easier if disappearance behind occluders is gradual rather than instant; (2) tracking is not influenced by edge type (visible vs. amodal); (3) tracking is compromised when objects change size before or after occlusion; (4) tracking is compromised when objects change motion direction behind an occluder (i.e. they reappear on the same side of the occluder) and (5) tracking is not compromised if all of the objects disappear behind the same occluder. In general then, it appears to be possible to track objects even when there is no continuous visual contact (Scholl & Pylyshyn, 1999; Scholl & Feigenson, 2004).

Slemmer and Johnson (2002) examined two more conditions involving occlusion. In the first condition, their occluders had blurred edges. Although participants could still track objects they were not as successful as in the Scholl and Pylyshyn study (1999). In the second condition trajectories were intersected and the objects occasionally occluded each other. Tracking under such conditions was also significantly harder.

Other researchers also examined various effects of occlusion on tracking. One of the experimental paradigms included simultaneous disappearance of all mobile objects. Alvarez, Horowitz, Arsenio, Dimase, & Wolfe (2005) found that participants were quite successful at tracking in this paradigm, treating the simultaneous disappearance as complete occlusion. Horowitz, Birnkrant, Fencsik, Tran, & Wolfe (2006) found that participants were more successful in the condition of simultaneous than asynchronous disappearance.

Keane and Pylyshyn (2006) concluded that object tracking behind occluders was so robust that there was no failure in performance even when the object was invisible for up to 900 ms. Interestingly, tracking accuracy did not decrease for objects moving behind occluders in an unpredictable way, a result contradicting the (4^{th}) conclusion from Scholl & Pylyshyn's previous study (1999).

Flombaum, Scholl & Pylyshyn (2008) were interested in the precise mechanism underlying object tracking behind occluders, proposing a special role for attention in this task. In their experiment, participants were asked to detect a dot, which „appeared sporadically on a target, a distractor, an occluder or in empty space" (Flombaum, Scholl, & Pylyshyn, 2008, page 904). The pattern of results strongly suggests that participants were attentively tracking objects even during the occlusion. That is, the percentage of correctly detected dots was

significantly higher if there was an object (either a target or a distractor) behind the selected occluder. They concluded that tracking objects behind occluders demanded allocatin of special attentional resources, even when the task seemed quite simple (Flombaum, Scholl, & Pylyshyn, 2008).

MOT in 3D scenes

As already noted, in natural viewing conditions, objects tend to move behind occluders, becoming at least partially invisible to the observer. This means that (in comparison to the occluder) the object must be further away from the observer.

Inspired by such reasoning, Viswanathan and Mingolla (Viswanathan & Mingolla, 1998; Viswanathan & Mingolla, 2002) tested object tracking when their trajectories were placed in different depth planes. In their experiment, the objects were moving on two parallel planar surfaces. Participants performed better in the depth condition than in the condition when all of the objects moved across a single surface. The authors concluded that, in order to make tracking easy, attention must have been equally distributed between the two parallel surfaces. The result was further supported with the finding that the short-term visual memory capacity becomes larger when objects were positioned in different depths (Xu & Nakayama, 2007).

These results could also be interpreted according to Pylyshyn's FINST theory (Pylyshyn & Storm, 1988). The distribution of attention across two surfaces would result in the creation of two separate groups of reference tokens, with every group connected to the objects from one of the surfaces. This distribution of FINSTs would lead to an increased capacity of the object's tracking. In fact, such results were obtained for right and left visual hemifields, demonstrating that this kind of division leads to successful tracking of twice as many targets (Alvarez & Cavanagh, 2005).

Although the experiment by Viswanathan and Mingolla (1998) was an important step in introducing more realistic conditions, it was still conducted on a flat 2D surface. It remains unclear how prominent the impression of 3D surfaces can become, one placed behind the other, give this display environment. To overcome this problem, two depth cues were added: binocular disparity and T-junction. These cues significantly improved tracking especially in the conditions in which objects' paths intersected (Viswanathan & Mingolla, 2002).

MOT with depth cues

The exploration of depth cues was introduced early in the study of vision (Locke, 1690) to justify the visual systems ability to reconstruct the third dimension from two-dimensional retinal images. Some depth cues, such as convergence and binocular disparity, require both eyes (binocular cues), and some can be assessed already with one eye (monocular cues). Monocular cues include motion parallax, perspective, relative size, familiar size, aerial perspective, texture gradient etc. (for a review, see Proffitt & Caudek, 2002). Pictorial depth cues are a subset of monocular cues, and can be reproduced on

static pictures (linear perspective, texture gradient, occlusion, shadows, relative size, contrast etc.).

Liu et al., (2005) compared the objects tracking in 2D and 3D scenes. Although the subjects performed well in both experiments, there was a slight advantage for 3D scenes. In this particular experiment, apart from pictorial depth cues (grid floor, wire frame and relative size), the movement of objects was also used to create appearance of three-dimensional space. The authors divided pictorial depth cues into two groups based on their contribution to 3D appearance: cues directly affecting object appearance (such as relative size) and cues affecting the overall scene layout (grid floor, wire frame). Further experiments showed that the tracking wasn't affected by the removal of cues affecting scene layout (wire frame and grid floor). However, additional distortion of the surface and reduction in cohesiveness of the 3D scene significantly reduced tracking accuracy. The researchers concluded that multiple object tracking is based on scene– and not retinal-coordinates (Liu et al., 2005).

However, the importance of three-dimensionality in tracking tasks has not been universally confirmed. Zelinsky and collaborators (Zelinsky, Neider, & Todor, 2007; Zelinsky & Neider, 2008) used sharks swimming through an underwater scene as stimuli, in an attempt to examine multiple object tracking in more realistic 3D environments. Accuracy in their 3D task was very similar to that in 2D tasks: with 1 to 3 targets, percentage of correctly identified targets was 92%, but with one additional target percentage dropped down to 78%. Furthermore, tracking accuracy was not influenced (1) when the intersection of trajectories was introduced, (2) motion terminated at different apparent depths.

MOT in the present study

In our research we used both relative size and contrast to directly change the appearance of objects on the screen and we also introduced a texture gradient, a cue that would affect scene layout, creating an overall impression of three-dimensionality.

It is possible that the lack of an advantage for the 3D condition is due to the details of their display. Zelinsky and collaborator's stimuli did move in virtually displayed underwater scene, but again this scene was presented on 2D monitor. Thus, they also had to introduce some of the depth cues to create an impression of 3D (decreasing the size of the sharks as they moved further, occlusion during motion, etc.). These depth cues were mostly cues directly affecting objects, without use of any cues affecting the scene layout. Hence, it is possible that the whole scene did not appear 3D. In order to avoid this problem, in our research we introduced both depth cues directly affecting object appearance as well as depth cues affecting the scene layout. These cues were separately introduced during the experiment, and we examined influence of each cue on the object tracking in 3D scene.

In comparison to the standard tracking procedure we had two key additions to our display: depth cues and occluders. Standard MOT paradigm involved

tracking of identical moving objects across a flat two-dimensional surface. Although at first such a procedure enabled researchers to observe the tracking in its pure and simple form, these conditions eventually did not reveal the true nature of the phenomenon as they did not very well resemble natural situations. When we track our children in the park, surrounding cars on the highway or football players in the field, these objects are rarely identical, permanently and fully visible, at the same depth, etc. Several studies already dealt with non-identical objects (Horowitz et al., 2007; Vidaković & Zdravković, 2009). In the present study we are concerned with the tracking of occasionally occluded objects in three-dimensional space. As in previous research, we included depth cues as well as occluders to create a sense of three-dimensionality on the computer screen. Though the positive effect of three-dimensionality was not generally confirmed (Liu et al., 2005; Zelinsky, Neider, & Todor, 2007; Zelinsky & Neider, 2008) we were still interested in the phenomenon in the complex natural-like conditions. Introduction of occluders adds to the complexity since the objects were not continuously visible.

Though the previous research introduced both depth cues and occluders, our research includes a set of depth cues not previously tested (separately or in combination). Also there were characteristics of the occluders not previously examined such as depth and number of occluders. We introduce a condition with occluders at different depth planes. In the previous research larger number of mobile occluders allowed each object to disappear behind its own occluder at any time and any screen position (Scholl & Pylyshyn, 1999). In contrast, we kept occluders at fixed position leaving every object equal opportunity to be occluded by any of them.

Method

Participants: A different group of twenty-two psychology students from the Faculty of Philosophy, University of Novi Sad, participated in each experiment. They were naïve to the purpose of the experiment and participated as a part of the course requirement. All of the observers had normal or corrected to normal vision.

Stimuli: Eight identical objects were presented on a white computer screen. When presented statically those objects appeared as gray circles, but as soon as they started to move they looked like rolling balls. Apart from moving, the objects changed their appearance in correspondence with certain depth cues. Depending on the experimental condition different depth cues were applied. The three depth cues, relative size, contrast and texture gradient, were applied either separately or in combination.

Relative size provides information about the relative depth of two objects, but only for familiar objects or objects of about same absolute size. The object that subtends the larger visual angle on the retina, appears closer to an observer (Proffitt & Caudek, 2002). In our display, the objects on the bottom of the screen had a diameter of 2° of visual angle, and those at the top of the screen 1° of visual angle. Hence objects appeared closer when presented on

the bottom than at the top of the screen, in accordance with the usual pictorial representation in western art. In this manner the two cues, size and the height in the visual field, present a congruent combination creating strong sense of depth. There is another cue used in the western art that contributes to this combination: horizon. Objects that are further away are closer to the horizon and hence „higher" in the visual field. In paintings this characteristics translates into the 2D position closer to the upper border of the painting for the objects represented in the further depth plane.

A change in contrast simulates the effects of aerial perspective, typical in natural conditions. An object that has larger contrast to the background appears closer to an observer (Proffitt & Caudek, 2002). Distant objects, due to the light scattering by atmosphere, appear less saturated in color, lowering the contrast with the background at the same time. In our displays, the objects' contrast decreased as they moved from the bottom of the screen to the top of the screen, consistent with the impression of reduction in color saturation, as the ball appeared to roll away in the depth. Objects were black (0% brightness) when presented on the bottom of the screen, and light gray (50 % brightness) on the top of the screen.

A texture gradient is a surface pattern that provides information about distance, depth and shape of the objects. An object, with larger and/or less dense texture elements, appears closer to an observer (Proffitt & Caudek, 2002). In our display we created „a floor" on which objects roll away into apparent depth and which itself contributed to the impression of three-dimensionality. The floor was trapezoid, representing a square's projection in depth, surrounded by „walls" on three sides, creating strong perspective cues (figures 1–5). The bottom edge of the trapezoid subtend 28° of visual angle, the top edge 14°, while its height subtend 12°. The floor was overlaid with 9x9 grid, which again followed the laws of perspective, adding to the impression of 3D space and also allowing to control starting positions and trajectories of the objects.

The three depth cues, relative size, contrast and texture gradient, were introduced separately or combined in order to examine the contribution of each cue and their relative importance.

Out of eight identical objects four were the targets in the tracking task and the rest were distractors. The number of targets was established based on the previous studies showing that tracking is optimal for 3–5 objects (Pylyshyn & Storm, 1988; Oksama & Hyönä, 2004).

Finally, occluders were added to the scene. Their number (2–4) and position varied depending on the experimental condition. In order to create the effect of depth, the height and width of occluders changed with their screen position but each occluder always occupied the same number of floor elements on the grid: one square in width and three squares in height. There were two occluder constellations: (1) they were all arranged in the same row, appearing to be at the same distance from observer, (2) They were scattered across the scene, which made them appear to be positioned in different depths. In either case they were not positioned to occlude each other, but each and every occluder was fully visible and perceived as a separate visual object.

The animation for this experiment was made in Adobe Flash software.

Procedure: Observers were placed 57 cm away from the 15" monitor. They were given the instruction, completed one exercise trial and proceed to the experimental task.

The objects, eight static circles on white background, were randomly placed on one of the 81 screen positions (on 9x9 floor grid) and shown to the observer. At the beginning of each trial four of the circles were marked as targets. Observes were supposed to track those

targets and to identify them at the end of each trial. After the observers understood which objects were the targets, all of the eight objects started to move, creating an impression of eight balls rolling around in 3D space. They were moving at the speed of 50 frames per second or 3.84 floor elements per second. The paths of the balls in 3D were designed so that there were no intersections (consequently the balls would never occlude each other). In the course of motion objects altered their contrast and size, depending on their position on the screen, and in accordance with the relevant depth cues.

At the end of the trial observes used the mouse to click on four objects, marking them red to indicate the position of the targets. They were allowed to change the selection before they moved to the next trial. There was no feedback.

Two measures were of interest (1) accuracy: percentage of correctly located targets and (2) reaction time: time needed to complete each trial (measured from the moments the targets stopped until all of them were marked by the observer). Although observers were told that we measure both reaction time and accuracy, they were not asked to proceed faster through the trials.

EXPERIMENT 1

Experimental design

In this experiment there were 12 trials in each of the five blocks (total 60 trials) /Picture 1/:

1) The first block contained trials that tested the texture gradient alone. The grid on the floor created the gradient and the appearance of 3D, but the objects themselves did not change (i.e. their relative size and contrast), as they appeared to roll away in this 3D space (Figure 1).
2) The second block contained the trials that tested contrast alone. The object would change contrast, as they appeared to roll in depth (Figure 2).
3) The third block contained the trials that tested relative size alone. The object would change their size, as they appeared to roll in depth (Figure 3).
4) The fourth block was one of the two control conditions. In this block all of the depth cues were simultaneously present (Figure 4).
5) Fifth block, the second control condition, contained trails in which none of the depth cues was present (Figure 5).

In each block the number and position of the occluders was balanced. Additionally the order of blocks was randomized for each observer and the order of 12 trials in each block was also randomized for each observer.

Results

The first goal was to establish weather the presence of different depth cues would produce separate effects on the tracking accuracy. One way ANOVA (repeated measures) showed that number of correctly located targets significantly[1] differ for different blocks ($F(4,96)=2.612$, $p=.040$).

1 In this study significance level $p<.05$ was applied

Figure 1

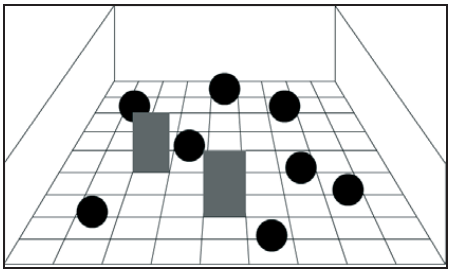


Figure 2

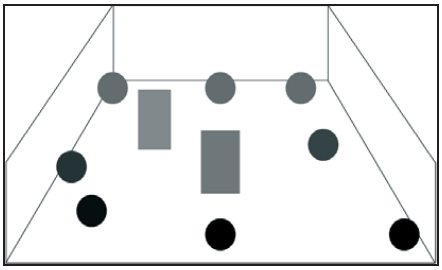


Figure 3

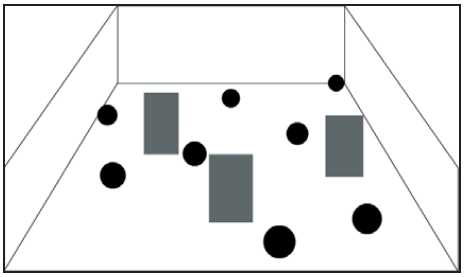


Figure 4

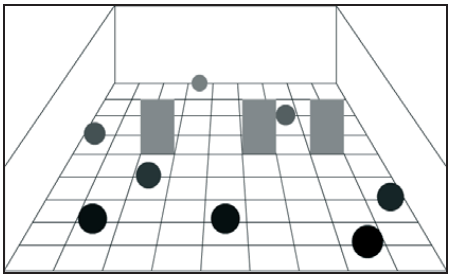
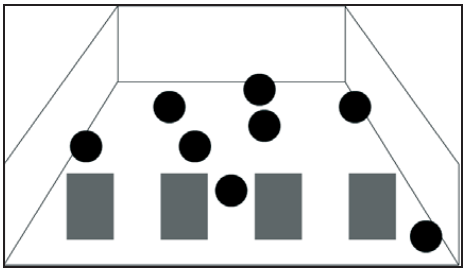


Figure 5



Picture 1. Five blocks in Experiment 1

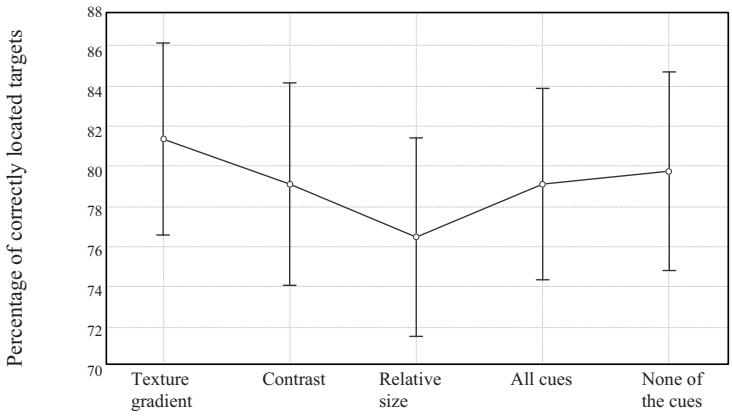
Further analysis (Fisher post-hoc test, Table 1) revealed the contribution of each depth cue. In comparison to the control conditions, the texture gradient did not reach significance, a result in agreement with work by Liu and collaborators' (Liu et al., 2005). In their experiments the texture gradient did not contribute to the perceptual cohesiveness of the 3D scene and consequently did not contribute to the tracking accuracy. However, we still included this cue because of the interpretation they offered at the end of the paper. Namely, despite their findings they concluded that the object tracking was based on the scene coordinates. This could only mean that the more complex scene should improve the object

tracking. We introduced such a complex scene. Still, it is possible that in our rich scene, the 3D motion and occluders scattered in depth already made a scene so coherent that the gradient did not lead to further measurable effects.

Table 1. Fisher post-hoc test

	texture gradient	contrast	relative size	all cues	none of the cues
Texture gradient					
Contrast	0.149				
Relative size	0.002*	0.090			
All cues	0.149	1.000	0.090		
None of the cues	0.301	0.679	0.036*	0.679	

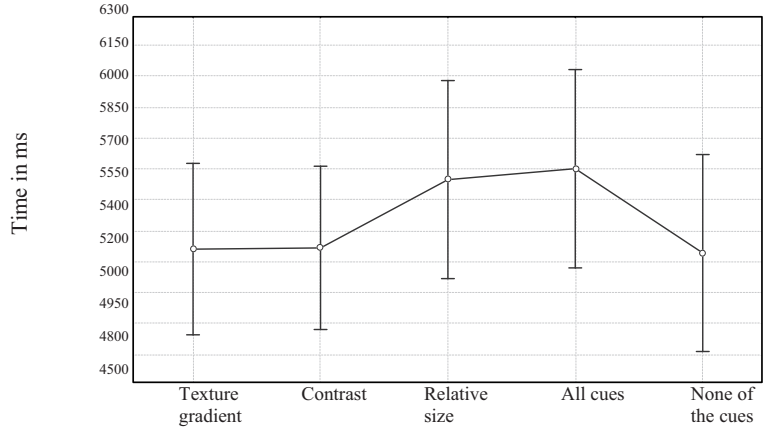
* Statistically significant differences



Graph 1. Percentage of correctly located targets within each block (vertical bars denote 0.95 confidence interval)

Graph 1 also shows that with varying object size participants were quite inaccurate in location judgment, implying that relative size is rather ineffective depth cue (in negative correlation with texture gradient). Additionally, there were no significant differences between the all cues condition and each of the three conditions with a single cue.

The results for the time needed to finish each trial were similar to the results obtained for the percent of correctly located targets (graph 2). One way ANOVA (repeated measures) showed significant difference between the trials ($F(4,96)=4.814$, $p=.001$).



Graph 2. Time spent for task finishing within each block
(vertical bars denote 0.95 confidence interval)

Fisher post-hoc test showed significant differences between the blocks (Table 2).

Table 2. Fisher post-hoc test

	texture gradient	contrast	relative size	all cues	none of the cues
Texture gradient					
Contrast	0.963				
Relative size	0.011*	0.013*			
All cues	0.004*	0.004*	0.686		
None of the cues	0.885	0.849	0.008*	0.002*	

* Statistically significant differences

Similar results were obtained for reaction times. Again the change of relative size decreased the objects' tracking accuracy. Thus, changing the relative size produced a smaller number of correctly located targets, as well as longer task time. The reaction time proved to be more sensitive independent variable than the number of correctly located targets showing a larger number of significant differences between the blocks.

In this experiment, we also examined influence of occluders, their number and position. Occluders would either completely or partially occlude the objects for a portion of time.

There were two possible positions for occluders: (1) all of the occluders were in the same row (appearing to be in the same depth), and (2) they were

scattered on the screen (appearing to be in the different depth planes). The results (one way ANOVA for repeated measures) showed that the participants were significantly more efficient in locating the targets when the occluders appeared to be in different depth planes ($F(1,24)=40.899$, $p=0.000$). However, there was no difference in the time needed to perform the task ($F(1,24)=.0396$, $p=0.844$). Since the accuracy was increased but the time remained the same, it could be proposed that the scattered occluders contributed to the cohesiveness of 3D scene, and thereby they made the object tracking easier. They might have been treated as additional depth cues.

There were 2, 3 or 4 occluders present in the scene. They made a significant effect on the objects tracking accuracy ($F(2,48)=13.586$, $p=.000$) and the time needed for the task ($F(2,48)=11.198$, $p=.000$). The presence of four occluders both decreased the accuracy and increased the time.

Apart from the described analysis, we run the ANOVA for three factors but none of the interactions reached significance. The first dependant variabe was number of corectly located targets: category and poisition of occluders ($F(4, 1470)=.19$, $p=.94$), category and number of occluders ($F(8, 1470)=.93$, $p=.36$), poisition and number of occluders ($F(2, 1470)=.67$, $p=.45$), interaction of the three factors ($F(8, 1470)=1.45$, $p=.17$). The same was obtained with the second dependant variable, time needed to complete task: category and poisition of occluders ($F(4, 1470)=.62$, $p=.65$), category and number of occluders ($F(8, 1470)=.60$, $p=.78$), poisition and number of occluders ($F(2, 1470)=.42$, $p=.66$), interaction of the three factors ($F(8, 1470)=1.22$, $p=.28$).

Finally, there is a small but significant negative correlation between the number of correct answers and task time ($r=–0.33$, $p=.000$). This correlation means that participants were slower in the tasks where they made more mistakes, suggesting that there was no speed-accuracy trade of but simply the tasks were becoming more difficult.

Discussion

The purpose of Experiment 1 was to examine whether the presence of different depth cues influenced object tracking in a 3D scene. We examined effects of three types of depth cue: texture gradient, relative size and contrast. Five blocks were constructed and these cues were separately varied. Differences in the object tracking accuracy were measured as a result of these variations. As a measurement of accuracy we took the number of correctly located targets and the time spent on each task. The results suggest that texture gradient, the depth cue indicating scene layout, did not have an effect on MOT task. Unfortunately, results did not clarify the difference between relative size and contrast.

Although we tried to ensure a strong impression of depth on our display, it was still presented on a 2D screen and it is possible to try and interpret results in this view. In this case, the size constancy would fail, rendering objects smaller for

a portion of time (i.e. the time they were suppose to look as if they were further away). It might be argued that it is more difficult to track smaller objects, which in our analysis produced relative size into an ineffective depth cue. However, we do not believe this to be the case, since this logic would need to be applicable to the other depth cue influencing appearance of the objects and that is contrast. The objects represented to be further in depth, had lower contrast against the background, which again would make tracking more difficult. However we did not measure ineffectiveness of contrast. Therefore we suggest that the observers responded to our stimuli layout as if it was a 3D scene.

Our purpose was to establish the contribution of particular depth cues in the MOT paradigm. However the separate contribution relative size and contrast was not established. The second experiment was designed to examine these differences.

EXPERIMENT 2

The purpose of the second experiment was to specify exact contribution of the two depth cues that influence the object appearance in the MOT task. Our goal was to keep the set up as similar to Experiment 1, while manipulating the cue contribution in a novel manner. To this end we introduced the manipulation, which included the incongruent change of the depth cues.

Method

Stimuli: The stimuli and the task were similar to those used in Experiment 1. Again we used the MOT paradigm with eight objects (four targets, four distractors) in a computer scene designed to appear 3D. However, since the purpose of this experiment was to examine the separate contribution of relative size and color perspective, new variations were introduced. We introduced a condition in which object alterations mimicking the influence of the different depth cues were conflicted. For example, a bigger object (2° of visual angle) that should appear closer to the observer would change its color to light gray (50 % brightness) as if it was in a far plane (i.e. the cues were changing in the opposite directions creating a perceptual conflict). Meanwhile a smaller object (1° of visual angle) would be black (0 % brightness). Hence size and color cues were contrasted allowing us to measure their importance and contribution within MOT paradigm.

There was no grid on the floor, which excluded the scene layout as a possible depth cue. The occluders were manipulated in the same manner as in Experiment 1.

Experimental design: There were 60 trials divided into five blocks (Picture 2). First two blocks tested each cue separately, third block was a control and the last two blocks contained conflicting depth cues.

1) In the first block, the contrast decreased as objects appeared to move in depth, and relative size remained unchanged (Figure 6).
2) In the second block, the size of the objects decreased as they appeared to move in depth, and contrast remained unchanged (Figure 7).
3) This block was a control condition and none of the depth cues were used (Figure 8).

4) In the forth block the contrast changed so that the objects became darker when they appear to recede in depth, but the size of the object changed in the opposite direction (Figure 9).
5) In the fifth block, the size of the object decreased when they appear to recede in depth, but contrast changed in the opposite direction (Figure 10).

Again, the blocks as well as trials within the blocks, were randomized across the obsevers.

Figure 6

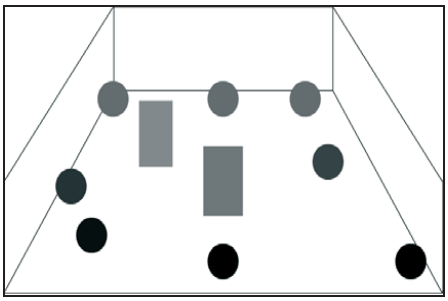


Figure 7

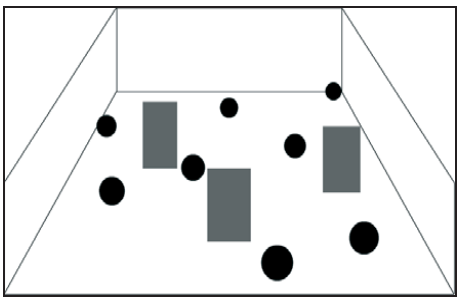


Figure 8

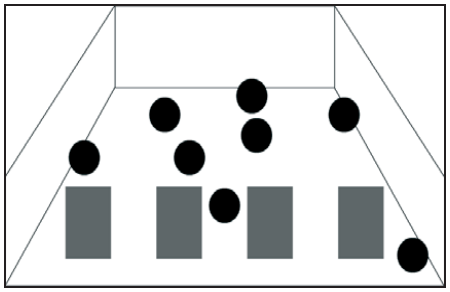


Figure 9

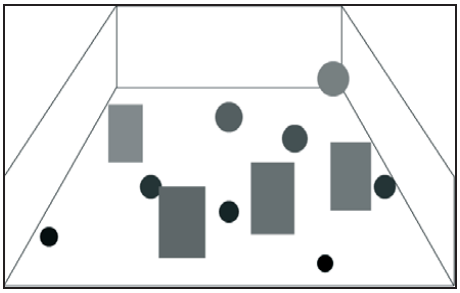
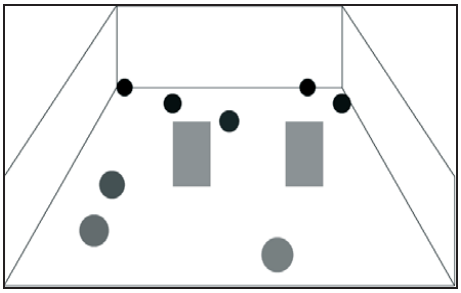


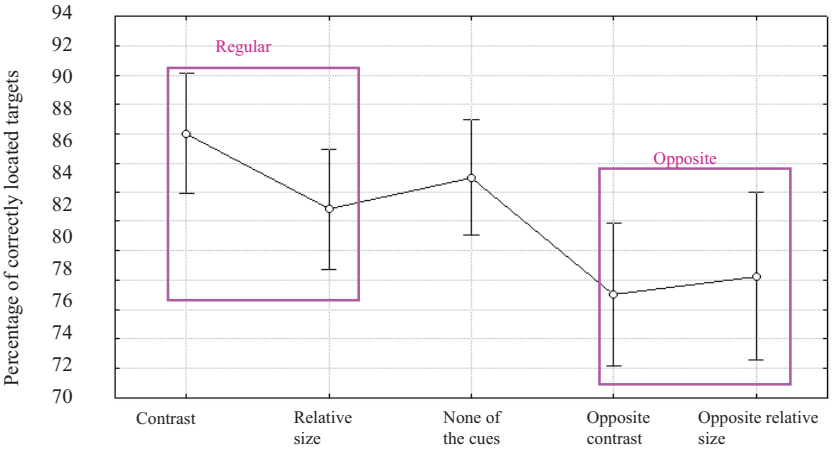
Figure 10



Picture 2. Five blocks in Experiment 2

Results and discussion

A one-way ANOVA (for repeated measures) showed that there was a significantly different number of correctly located targets between the blocks (F(4,84)=12.872, p=.000), Graph 3.



Graph 3 – Percentage of correctly located targets within each block (vertical bars denote 0.95 confidence interval)

Fisher post-hoc test showed significant differences between the conditions (Table 3):

Table 3. Fisher post-hoc test

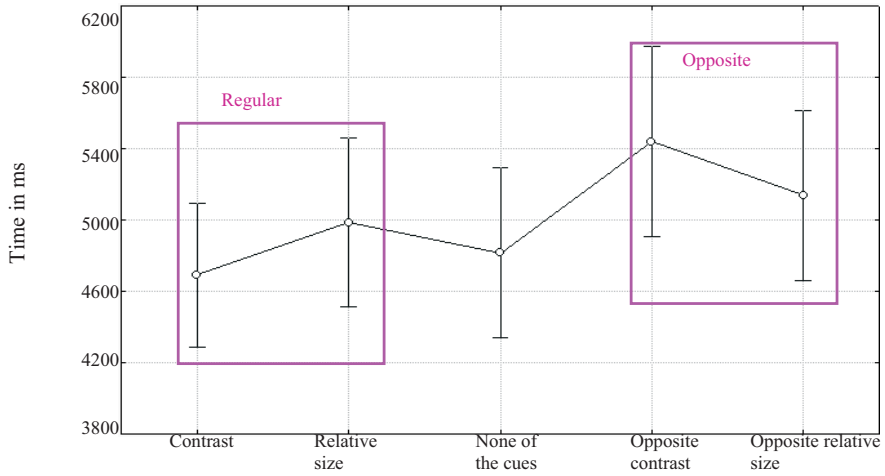
	contrast	relative size	all cues	none of the cues	opposite contrast
Contrast					
Relative size	0.005*				
None of the cues	0.101	0.231			
Opposite relative size	0.000*	0.014*	0.000*		
Opposite contrast	0.000*	0.002*	0.000*	0.484	

* Statistically significant differences

As can be seen on the graph 3, there is a notable difference in performance in the conditions with a single cue („regular“) and the conditions in which that cue is obstructed by another cue („opposite“). Observers are significantly worse when the cues change in a conflicting manner.

However the question posed in this experiment was whether the contrast and the relative size produce different effects on object tracking. Our results demonstrate that there is a measurable difference in the influence of the two cues and that contrast has a stronger effect on object tracking in 3D space. However these two situations are not significantly different from the control which limits the conclusions.

This is additionally supported by the results obtained for the time it took observers to go through the trials. Here they need significantly more time for the conflicting conditions ($F(4,84)=14.399$, $p=.000$).



Graph 4. Time spent for task finishing within each block
(vertical bars denote 0.95 confidence interval)

Fisher post-hoc test showed significant differences between the blocks (Table 4).

Table 4. Fisher post-hoc test

	contrast	relative size	all cues	none of the cues	opposite contrast
Contrast					
Relative size	0.008*				
None of the cues	0.253	0.122			
Opposite relative size	0.000*	0.166	0.004*		
Opposite contrast	0.000*	0.000*	0.000*	0.007*	

* Statistically significant differences

Reaction times were in concordance with the number of correctly located targets. Again, there is a negative correlation between the number of correct answers and the reaction time (r=–0.476, p=.000), indicating that more time was devoted to the tasks with less correct answers. Conflicting conditions needed both longer time and produced more errors, suggesting that they were significantly more difficult for the observers. Direct comparison between the two situations in which a single cue is used, shows that contrast produces faster responses. On the other hand, comparing the situations with conflicting cues we can see that the irregular change of contrast led to slower responses.

Taken together, these findings indicate that the contrast might be more important cue for depth perception in general. If presented in a simple task, it helps both speed and accuracy, but if presented in a conflicting task it damages the objects' tracking. This finding is in concordance with O'Shea and collaborators' results (O'Shea, Blackburn, & Ono, 1994). Though their experimental task was different they still showed contrast to be more significant than relative size, contributing more to the perceived depth even when conflicting with relative size.

The presence of occluders produced the same effects as in Experiment 1. The number of correctly identified targets was significantly higher in the condition with scattered occluders ($F(1,21)=11.047$, $p=.003$), but there was no difference in reaction time. Also just like in Experiment 1 the larger number of occluders led to less accuracy ($F(2,42)=13.052$, $p=.000$), and longer reaction time ($F(2,42)=3.629$, $p=.035$).

Again there was no interaction between the factors (ANOVA, repeated measures). The first dependant variabe was number of corectly located targets: category and poisition of occluders ($F(4, 1290)=1.79$, $p=.13$), category and number of occluders ($F(8, 1290)=1.43$, $p=.18$), poisition and number of occluders ($F(2, 1290)=.48$, $p=.62$), interaction of the three factors ($F(8, 1290)=1.69$, $p=.09$). The same was obtained with the second dependant variable, time needed to complete task: category and poisition of occluders ($F(4, 1290)=.78$, $p=.52$), category and number of occluders ($F(8, 1290)=1.2$, $p=.29$), poisition and number of occluders ($F(2, 1290)=.59$, $p=.55$), interaction of the three factors ($F(8, 1290)=.97$, $p=.46$)

The purpose of this experiment was to examine differences in influence of the two pictorial cues, which directly affect object appearance. The two cues proved to contribute unequally to the object tracking in apparent 3D space. Contrast seems to have a larger influence.

DISCUSSION AND CONCLUSION

Tracking moving objects is a typical everyday task and the tracked objects could be children, cars or teammates. These objects seldom move in the same depth plane (i.e. keeping the same distance from the observer). On the contrary, objects move closer or away from the observer. They also move in other directions (left-right, up-down). Therefore, it should be very important to

accurately estimate direction of movement in a number of every day situations (for example: is this car approaching).

How do people solve such a task? Our initial assumption was that the depth cues help in the estimation of motion direction, when motion occurs in 3D space. However, not all of the depth cues have the same status: some define scene layout and some contribute to object appearance. Also some of the cues could be absent from certain scenes. Do these differences affect performance in multiple object tracking task?

Three depth cues were examined in Experiment 1: texture gradient, relative size and contrast. The texture gradient, cue attributing to scene layout, did not seem to contribute to the object tracking in three-dimensional space. This finding is consistent Liu et al., (2005), who found no decrease in object tracking accuracy when texture gradient was removed. Paradoxically they still concluded that the object tracking was based on scene layout and that decreasing a number of details in the scene should lead to poorer tracking performance. The lack of the effect in their own study was explained by general ineffectiveness of texture gradient in that particular scene layout. The texture gradient simply did not contribute to scene three-dimensionality consequently its removal did not influence the performance. In our rich scene it was also ineffective.

The other two depth cues we have investigated were related to object appearance. The change in relative size produced a negative effect on tracking accuracy. Observers both produced more errors and were slower when tracking object undergoing size change during their motion. On the other hand, there was not such an effect for the other investigated cue, contrast. The subjects were significantly more accurate when they made judgments based on contrast alone. Therefore we conducted the second experiment in order to broaden our understanding of the difference between the two cues.

Experiment 2 was designed to further investigate the difference between the effect that relative size and contrast produce on perception of motion in apparent 3D scenes. Influence of these two cues was not only measured but they were also placed into direct conflict with each other. That is, in some of the conditions the two cues were varied in „opposite“ directions sending contradictory information about objects' motion direction. The direct comparison demonstrates a larger effect of contrast on objects tracking. When it is congruent with the rest of the scene, contrast significantly helps, while in incongruent condition contrast harms speed and accuracy of the performance.

The MOT paradigm is not the only experimental task that demonstrates the supremacy of contrast. O'Shea, Blackburn, & Ono (1994) also emphasized the importance of contrast. In their research, contrast proved to be an important and reliable cue, trusted by participants even when relative size produced conflicting depth information. Our previous research also offers similar findings (Vidaković & Zdravković, 2009). We found that color is a most significant

characteristic in a multiple-object-tracking task, performed in two-dimensional space. Consequently contrast should also play an important role.

In the present study, occluding objects were also included in an attempt to make the tracking task as ecologically valid as possible. A larger number of occluders led to poor performance suggesting that the additional objects crowed the scene. Scholl and Pylyshyn (1999) however, proposed that reference tokens (FINSTs) could subsist even in the case of occlusion. Consequently, the decreased accuracy in the presence of larger number of occluders implies that connection between FINSTs and objects was not in fact so strong. Additionally, constant occlusion frequently interrupts these connections. The decreased performance probably signals the cognitive system's inability to continue tracking while being repeatedly interrupted. Our results are supported by Slemmer and Johnson (2002), who not only used specific occluding objects but also had their tracked objects occluding each other as they moved. The tracking was much harder under these conditions and authors theorized that the participants could hardly individualized objects. Slemmer and Johnson considered their experimental conditions to be ecologically valid unlike typical tracking task.

Here we have tried to go even further in creating ecologically valid conditions. Our occluders had different positions in the scene, adding to the appearance of depth in some conditions. As we predicted, the tracking accuracy was increased in these conditions with occluders at different depths. We believe that the efficiency of our observer was enhanced because the occluder's position served as yet another depth clue adding to the scene cohesiveness. Our conclusion is motivated by the Liu and collaborators findings (Liu et al., 2005) that object tracking was based on scene coordinates.

We also noticed that there was no speed-accuracy trade-off. The more difficult tasks simply render both more errors and longer solving time.

Finally the most important finding is that some depth cues contribute more than the others. Previous research showed that the situation (e.g. fixed objects distance estimation, tracking mobile objects, observer in motion) dictates which cues will have more weight. Surdick, Davis, King, & Hodges (1997) demonstrated that perspective cues (linear perspective, texture gradient) were the most important in distance estimation. However, in our experiments with a tracking task, the texture gradient did not play any role. We proposed that as our task required a shift of attention from scene to objects, the cues relevant for objects became more significant. In that respect, the contrast proved to be most relevant cue in our task. Relative size was also tested but did not proved to be as compelling when pitted against contrast.

Everyday situations require tracking of moving objects in three-dimensional space full of other objects (occluders and distractors). We tried to bring these conditions in the lab in order to study the important mechanisms of object' tracking. We found that contrast was most helpful cue while the large number

of occluders decreased the efficiency. When the occluders were also scattered in different depth planes, it helped with tracking. However, when the scene is too cluttered making the task complex, observers are slow and performance drops.

Acknowledgments. Authors would like to thank Ian M. Thornton and Oliver Tošković for valuable comments on the manuscript.

REFERENCES

Alvarez, G. A., & Cavanagh, P. (2005) Independent resources for attentional tracking in the left and right visual fields. *Psychological Science*, *16*(8), 637–643.

Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., Dimase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology-Human Perception and Performance, 31*(4), 643–667.

Flombaum, J. I., Scholl, B.J. & Pylyshyn, Z. W. (2008). Attentional resources in visual tracking through occlusion: The high-beams effect. *Cognition*, *107*, 904–931.

Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin & Review*, *13*, 516–523.

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics*, *69*(2), 172–184.

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology*, *52*, 346–368.

Kellman, P. J., & Shipeley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, *23*(2), 141–221.

Liu, G., Austen, E. L., Booth, K. S., Fisher, B., Rempel, M. I., & Enns, J. T. (2005). Multiple object tracking is based on scene not retinal coordinates. *Journal of Experimental Psychology: Human Perception & Performance*, 31, 235–247.

Locke, J. (1690). *An Essay Concerning Human Understanding*. Oxford: Clarendon Press, 1975.

O'Shea, R., Blackburn, S. G., & Ono, H. (1994). Contrast as a depth cue. *Vision Research*, *34*, 1595–1604.

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition*, *11*(5), 631–671.

Proffitt, D. R., & Caudek, C. (2002). Depth perception and perception of events. In A. F. Healy and R. W. Proctor (Eds.), *Comprehensive handbook of psychology, Volume 4: Experimental psychology (*pp. 213–236*)*. NY: Wiley.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.

Scholl, B. J., & Feigenson, L. (2004). When Out of Sight is Out of Mind: Perceiving Object Persistence Through Occlusion vs. Implosion [Abstract]. *Journal of Vision*, *4*(8):26, 26a.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking Multiple Items Through Occlusion: Clues to Visual Objecthood. *Cognitive Psychology*, *38*, 259–290.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*, 159–177.

Slemmer, J. A., & Johnson, S. P. (2002). Object tracking in ecologically valid occlusion events [Abstract]. *Journal of Vision*, *2*(7):239.

Surdick, R. T., Davis, E. T., King, R. A., & Hodges, L. F. (1997). The Perception of Distance in Simulated Visual Displays: A Comparison of the Effectiveness and Accuracy of Multiple Depth Cues Across Viewing Distances. *Presence*, *6*(5), 513–531.

Vidaković, V., & Zdravković, S. (2009). Color Influences Identification of the Moving Objects More Than Shape. *Psihologija, 42* (1), 79–93.

Viswanathan, L., & Mingolla, E. (2002). Dynamics of Attention in Depth: Evidence From Multi-Element Tracking. *Perception*, *31*, 1415–1437.

Viswanathan, L., & Mingolla, E. (1998). *Attention in depth: Disparity and occlusion cues facilitate multi-element visual tracking* (Tech. Rep. No. CAS/CNS–98–012). Boston: Boston University Center for Adaptive Systems, Department of Cognitive and Neural Systems.

Xu, Y., & Nakayama, K. (2007). Visual Short-Term Memory Benefit for Objects on Different 3-D Surfaces. *Journal of Experimental Psychology: General*, *136*(4), 653–662.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340.

Zelinsky, G., & Neider, M. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, 16 (5), 553–566.

Zelinsky, G., Neider, M., & Todor, A. (2007). Multi-object tracking in a realistic 3D environment [Abstract]. *Journal of Vision*, *7*(9):895, 895a.