On the failure of stereomotion capture in an object disappearance paradigm

Lora T. Likova[a, b] and Christopher W. Tyler[a]

[a]The Smith-Kettlewell Eye Research Institute,
2318 Fillmore St, San Francisco, CA 94 115
[b] Bulgarian Academy of Sciences, Sofia, BG 1113
lora@ski.org, cwt@ski.org,
www.ski.org/cwt

ABSTRACT

Motion capture is one of the basic effects of a moving surround. To explore the existence of a motion capture in the stereodomain, we designed dynamic autostereograms with target-surround configuration. The images consisted of several horizontal lines of disparate disks with the central line of disks specified as a target. The surround was set in stereomotion by alternating between two disparities, while the target did not change in disparity, but suddenly disappeared in phase with one of the two surround depth-planes, reappearing in phase with the other surround plane. The target was vividly perceived as moving in depth despite its lack of any disparity change or even a paired location in the second of the two alternating frames. The motion capture hypothesis predicts that the target should be seen to move in synchrony and in the same direction as the surround. However, surprisingly, the data showed that the target was always perceived disappearing in a backward direction and reappearing in a forward direction irrespective of the surround direction, thus suggesting that the reported illusory depth-motion into a stationary target is an independent perceptual phenomenon that has no relation to the expected capture paradigm.

Keywords: stereomotion, contextual effects, motion capture, object disappearance, motion interpretation, perceptual heuristics

1. INTRODUCTION

In the real world objects do not move in isolation, but within a complex environment that may have a strong influence on motion perception. Human behavior, from walking and driving to complex social behavior, depends very much on perceiving visual motion of objects and people surrounding us. Our perception can be strongly influenced by subtle factors in that global visual context, although motion studies usually do not take into a consideration the visual context in which motion is perceived.

This study is focused on the interpretative perceptual processing that is invoked in situations beyond the flat two-dimensional screen. Perception is an active interaction between the information stream entering the senses and the attempt of the higher brain areas to interpret what is "out there". The bottom-up and top-down streams interact between each other in a complex, poorly understood manner. We have approached the complexity of the real world through the perception of spatiotemporal relationships between a transient stereotarget and its global stereoscopic surround.

Motion capture (Ramachandran & Anstis, 1983; Ramachandran & Inada, 1985; Ramachandran & Cavanagh, 1985) is known to be one of the basic effects of a moving surround on a dynamic target: the moving surround may "capture" the target so that the target is perceived to move together with the surround. The motion capture phenomenon is so strong, that it is able to overcome and capture not only an incoherent dynamic noise, but even correlated random dot patterns moving in the opposite direction. Ramachandran & Cavanagh (1987) supposed that "motion capture suggests an important biological role for long-range apparent motion: to preserve the object identity while at the same time eliminating spurious motion signals arising from finer image features. In this manner the visual system solves the 'correspondence problem' for motion."

While well-studied in the frontoparallel plane, until now motion capture has not been investigated in the stereoscopic domain. To explore the existence of a stereomotion capture, we designed dynamic autostereograms with a target-surround configuration (Figure 1).

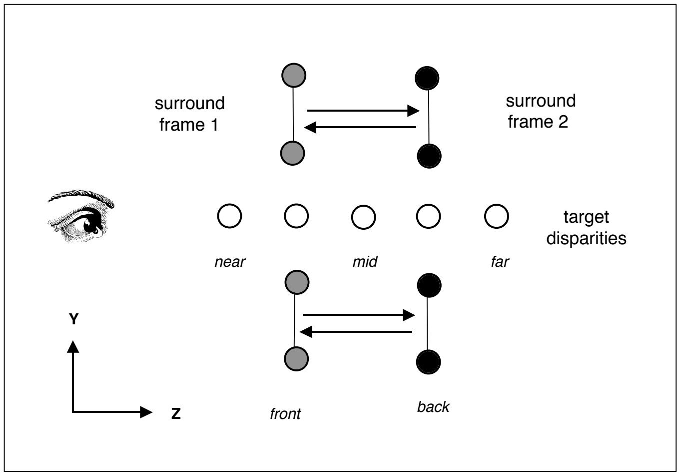

Figure1. Side view of the four TSD-configurations: *near TSD* – when the target (open circles) is before the surround (filled circles), *front* TSD– the target is in the front disparity plane of the surround, *mid* TSD - the target is between the front and back surround planes, *back* TSD - the target is in the back surround plane, *far TSD* - the target is behind the back surround plane.

The surround was set in stereomotion by alternating between two disparities, while the target did not change in disparity, but suddenly disappeared in phase with one of the two surround depth planes, and again reappeared in phase with the other surround plane. Two experiments were performed with different temporal relationships between the stereomoving surround and the appearing/disappearing target. In the first experiment, the target appeared in synchrony with the *near* surround plane (*front-synchronization*) and disappeared during presentation of the *far* surround plane. In the second experiment, the phasing was reversed (*back-synchronization*). Recent studies on motion (He & Nakayama, 1994; Nakayama et al., 1995) suggest that apparent motion correspondence and motion coding are determined at a level of *surface representation*, rather than on image shape or position. Where there is a gap in the information specifying the surface, this conceptualization suggests that the gap should be filled by interpolation of the defined surface into the empty region. Does this conceptualization apply to the front-synchronized and back-synchronized depth motions?

The operation of a stereomotion capture would predict in both cases the target to be perceived moving in the same depth direction and with the same magnitude as its surround. Consequently, the target should switch its direction of depth motion with the switch from front-synchronization to back-synchronization. This is the main issue evaluated in the present study.

2. METHODS

2.1. Stimuli

We used dynamic autostereograms consisting of several horizontal lines of disparate disks to generate stereomotion (Minev & Likova, 1999). The central line of disks specified as a target. The surround was set in stereomotion by alternating between two disparities (196 ± 6.5 arc min), while the target did not change in disparity, but suddenly disappeared in phase with one of the two surround depth planes, and again reappeared in phase with the other surround plane. The target disparity was fixed in five steps of 6.5 arc min around the center disparity of 196 arc min. The diameter of the disks was 26 arc min with a luminance of 52 cd/m^2 against a background of 0.31 cd/m^2. Five configurations of target/surround disparities (*TSD-configurations*) were explored: *near, front, mid, back, far* (see Fig. 1), where the egocentric target distance was 98.4 cm, 101.6 cm, 108.6 cm, 112.3 cm, respectively.

Vertical distance between the surround rows was 65 min. The principal presentation time was 600 msec for each frame, but the phenomena described were not restricted to this frame duration. The autostereographic display was viewed by free-fusion with uncrossed vergence of the eyes. To assess the role of vergence tracking in generating the stereomotion percepts, the task was performed using controlled fixation under two different conditions (i) target fixation and (ii) surround fixation on the row just above the target. If depth motion perception was mainly based on vergence movements, perceived target motion should be minimal with target fixation and much increased with fixation on the surround. If, on the other hand, depth motion perception was mainly based on contextual interactions within a global depth representation, the data should be essentially independent of fixation position.

2.2. Task

The observers' task was to provide estimates of (i) the direction and (ii) the magnitudes of both the target and surround motions. Two repetitions of the measures were performed for every condition.

2.3. Observers

Five observers (three male and two female, aged 22 - 58) with normal or corrected to normal vision took part in both experiments. The stimulus patterns were viewed at a 70 cm viewing distance.

3. EXPERIMENTAL RESULTS

Experiment 1: Front-synchronization

In the first experiment, the target appeared in phase with the front plane and its disappearance was in phase with the *back* plane of the stereomoving surround (*front-synchronization*). Synchronization of the target disappearance with the backward motion of the surround might be expected to *capture* the target so as to be perceived moving in congruence with the surround (analogous with the *lateral motion capture*). At the first glance, this prediction seemed correct: The observers reported a strong sense that the target moved backward in the *same* direction as the surround. Had we found a *motion capture* within the stereodomain?

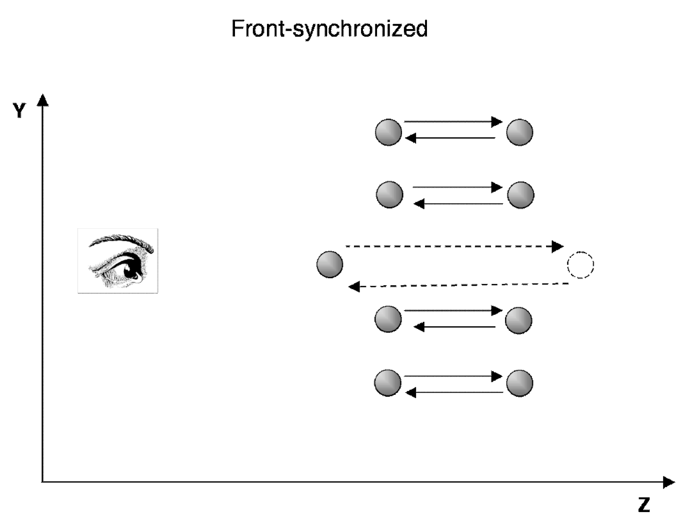

Figure 2. A schematic view of the motion percept, when a stereomotion was introduced into the surround. The target appears in synchrony with the front disparity plane of the surround, and disappears in synchrony with the further surround plane. The target was perceived moving backward when it disappeared (and forward when it again reappeared) and thus its direction was the same as the surround direction.

The basic prediction is that, if the motion of the target was *captured* by the motion of the surround, the perceived target motion would be expected to remain constant for all target disparities. If, on the other hand, the capture took the form of Gogel's (1972) *adjacency principle*, the capture would be expected to be strongest when the target was adjacent in disparity and synchronous with the surround. In this case, the magnitude of the perceived depth motion would be maximal for the *front -TSD* configuration and reduced for the other four disparity configurations. However, the results do not conform to either prediction.

Results

Figure 4 shows perceived motion magnitudes, averaged for five observers, against the TSD-configuration. The left and right panels show the data separately for fixation on the target and for fixation on the surround, respectively. Although a constant amount of physical stereomotion (7 cm) was present in the surround, the perceived stereomotion was not constant. Moreover, the target, despite being stationary, and even disappearing in every second frame, was experienced as moving backward, then forward again. Thus, the data show a striking deviation from the Stereomotion Capture Hypothesis and also deviate from the Depth Adjacency Hypothesis for near disparities. The trend for the target motion to decrease with distance from the observer (from TSD-*near* to TSD-*far)* was significant ($R^2 = 0.952$, $p = 0.003 < 0.01$) with target fixation and marginally significant ($R^2 = 0.90$, $p = 0.025 < 0.05$ with surround fixation).

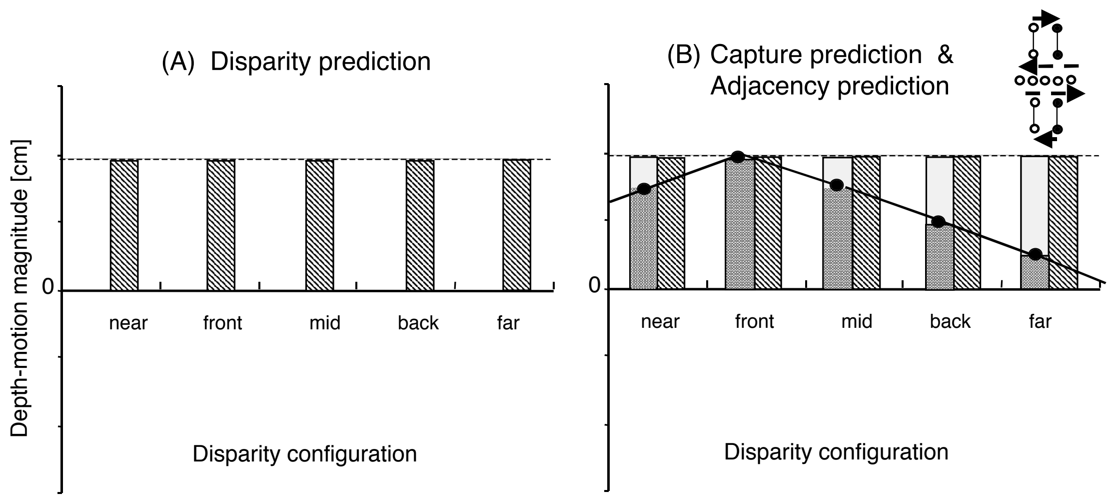

Figure 3. Predictions of A: the Disparity-change hypothesis and B: the Stereomotion Capture and Depth Adjacency hypotheses for the single-phase depth motion. The Disparity-change prediction (dashed line) is that only the surround should move with a constant magnitude regardless of target disparity (horizontal axis). The Capture hypothesis (open bars) predicts a fixed amplitude of capture relative to the surround location. The Adjacency hypothesis (full line) predicts that the target will be most affected when it is close in disparity to the surround at the time when both are presented.

Remarkably, not only did the illusory target motion resist any potential capture influence, it was even strong enough to suppress dramatically the disparity-specified surround motion. The reported stereomotion percept in the surround decreased in reciprocal fashion with the strength of the illusory target motion. This is a new effect of 'reversed' suppression of the *disparity*-defined stereomotion in the surround by an *illusory* motion in the target. It is particularly remarkable since (i) the 'pure' motion, emerging without any physical displacement, is able to affect and even dominate the 'real' disparity-defined motion and, moreover, (ii) to override the decisive role of the surround in a target/surround relationships.

This unexpected contextual suppression, as well the magnitude discrepancy, suggested that the observed *front-synchronized depth motion* was not governed by the motion capture mechanism, but by an independent generator deriving from the target events. The trend for the surround motion to increase showed that it was *also* significantly affected ($R^2 = 096301$, $p = 0.003 < 0.01$ under target fixation; $R^2 = 0.9329$, $p = 0.0075 < 0.01$ under surround fixation). The data were analyzed by Wilcoxon signed-rank test. The illusory motion magnitudes under both fixation conditions were not significantly different at $p > 0.05$ (N.S; $z = -1.752$; $p = 0.0796$).

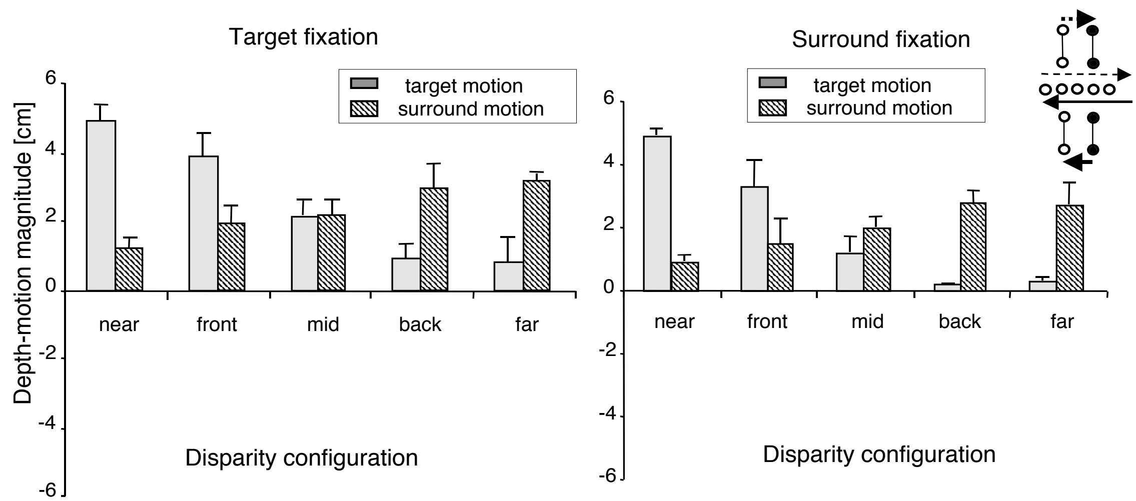

Figure 4. Perceived motion magnitudes of the target (gray bars) and surround (hatched bars) for five TSD-configurations. Error bars show 1 s.e.m. The front-synchronized depth motion under the two fixation conditions was not significantly different. The TSD-configurations are schematically depicted in the inset (see also Fig. 1). A strong sense of a moving target was experienced in the same direction as the surround. In fact, the target motion was almost unaffected by the presence of a stereomotion in the surround. Moreover, the illusory target motion had an unexpected *reciprocal effect* on the perceived motion magnitude in the surround.

Experiment 2: Back-synchronization: a strong test for stereomotion capture

To test conclusively the stereomotion capture hypotheses versus the more radical hypothesis of a new distinct mechanism as well as to explore further the interactions of an event of object disappearance, a second experiment employed the target disappearance in phase with the *near* disparity plane of the stereomoving surround (*back-synchronization*). Now the motion capture hypothesis predicts that the depth motion should also invert its direction.

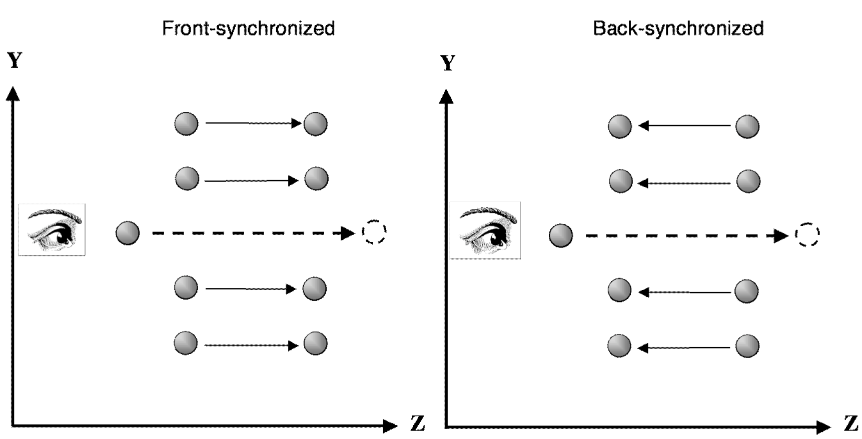

Figure 5. Schematic view of the perceived motion under the back-synchronized condition in comparison with the front-synchronized condition. In both cases the target was perceived moving backward when it disappeared, and forward when it reappeared again, in spite of the reversal in the direction of surround motion.

Results

Contrary to the motion capture prediction, the target did not switch its direction, but instead it was reported to move in the *opposite* direction to the surround motion. Figure 5 schematically presents the resulting percept in comparison with the percept under front-synchronized condition.

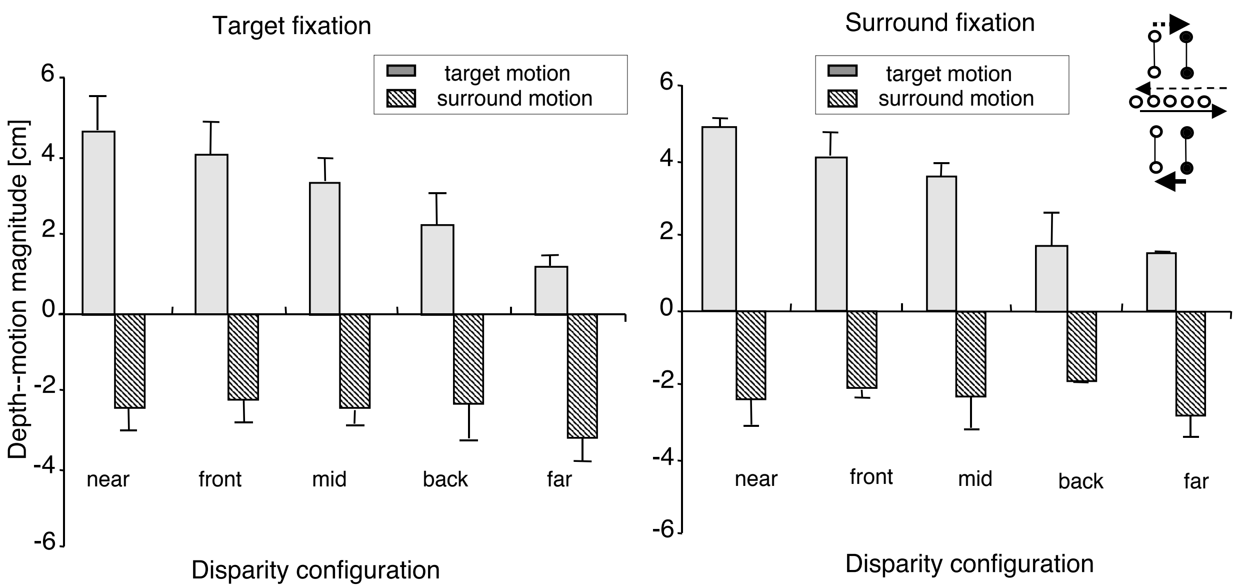

Figure 6. Perceived motion magnitudes of the target (gray bars) and surround (hatched bars) at the same five TSD-configurations as in the previous experiment were used (see inset and Fig. 2), but here the target appearance was synchronized with the 'appearance' of the *back* disparity plane. Error bars show 1 s.e.m. The back-synchronized depth motion was not significantly different under the two fixation conditions. The inversion of the surround stereomotion neither inverts the target direction, nor significantly influences its magnitude.

The averaged results for five observers are given in Fig. 6. The target was not changed in disparity and there was no other depth-motion cue, so no motion should be perceived in the target. However, as in Experiment 1, a profound depth motion was experienced in the stationary target. The back-synchronized motion magnitudes under both fixation conditions were not significantly different (N.S; Z = -0.73; $p = 0.4652, > 0.05$). The linear trend for the *target* motion to decrease with distance from the observer (from TSD-near to TSD-far) was significant under both fixation conditions ($R^2 = 0.986$, $p = 0.0007, < 0.01$ under target fixation; $R^2 = 0.947$, $p = 0.005, < 0.01$ under surround fixation).

In contrast, the *surround* motions were *not significantly* affected by the fixation modes in Experiment 2 ($R^2 = 0.484$, $p = 0.19 > 0.05$ under target fixation; $R^2 = 0.071$, $p = 0.0662 > 0.05$ under surround fixation).

If motion capture was working here, the target now should be perceived to *disappear* as it goes *forward* to the observer, in order to follow the surround. Nevertheless, the *back-synchronized* motion was reported to move opposite to the surround (Fig. 6). Evidently, the back-synchronized depth motion results do not support the *stereomotion-capture* hypothesis, because the surround was not able to switch the direction of perceived target as predicted by motion capture (i.e., the target did not disappear with a forward motion). Moreover, the perceived target motion seems to be independent from the dynamic behavior of the surround, always disappearing in a backward direction.

It is possible that the depth motion could be governed by the other basic effect of a moving surround on a stationary target: *motion induction.* Motion induction is a phenomenon of attributing motion to an object that had no moving counterpart in the retinal pattern. In the case when an object may be classified as a surround relative to another object, or when it acts as a frame of reference while the surround is in motion, the surrounded, enclosed object is perceived as moving in a direction opposite to its frame. Thus, the surround and the target motion are perceived in opposite directions. However, the data could not be explained by motion induction either, because if this was the mechanism, it should again apply in both *front-synchronized* and *back-synchronized* conditions. Thus, the stereomotion induction hypothesis implies opposite directions of depth motion for the front- and back-synchronized conditions, which is not the case (see Fig. 6).

4. DISCUSSION

To explore the existence of motion capture in the stereoscopic domain, we designed a dynamic stereo-configuration consisting of a transient target placed in a stereomoving surround. The target had no change in disparity, or even a paired event in the second of the two alternating frames. Surprisingly, in the three-dimensional context, a vivid sense of strong backward/forward motion was experienced as the target disappeared/reappeared. The pattern of results shows that none of the known effects of a moving surround could underlie the observed phenomenon.

On one hand, it might seems reasonable not to expect target motion to be perceived because there is no position change or motion cue presented in the target stimuli. However, an alternative point of view is also possible. It might be supposed that, in the absence of visual cues for motion in depth, or any kind of apparent motion, the brain would activate interpretive processes in its attempts to "make sense" of the events occurring in the world. The percept of the object dynamics would be determined by this interpretive level of processing.

Object Disappearance Effect – an interpretative mechanism
The comparison between front-synchronized and back-synchronized motions makes it clear that the target did not invert its direction in spite of switching in the surround motion. It follows that (1) both motions are actually governed by a common mechanism, which (2) differs from any of the available hypotheses (stereomotion capture, stereomotion induction, surface interpolation or depth adjacency) because if any of these were the mechanism, they should apply oppositely in the front-synchronized and back-synchronized conditions, which was not the case.

In order for events to be understood, some further processing is required, even after retinal images are organized into coherent objects and motions (Palmer, 1999). Moreover, the visual system is 'equipped' to 'perceive' or recognize sequences of events and their causal relationships (Michotte, 1963). Recognition in the case of a scene category refers to the classification of a scene into a semantic group (e.g., house, lake, forest, etc.). It has been established, for example, that adding an estimation of the mean *depth* of a stationary scene to other attributes may significantly increase performance through semantic recognition. The mean depth allows the emergence of specific semantic categories even if the groups overlap (Torralba & Oliva, 2002).

How do our brains understand the event of the target switching on and off? Some of the neural events elicited by the pattern of retinal motion are directed to classifying the object as a member of some known category (Palmer, 1999). The fact of a target switching on and off may be classified as the physical *event* of an object appearing and disappearing. When an object rapidly disappears from view, the event requires some cognitively consistent explanation. How could an object disappear? It could be occluded, it could be destroyed, or it could disappear from its last known position by moving somewhere. However, if the object moved forward or sideways, we would still see it in the visual field. Since there is no object visible in the second frame, the simplest explanation for object disappearance in a three-dimensional context is the heuristic assumption of a motion into the distance. Our data indicate that the surrounding stereomotion events actually have only a minor effect on the target dots. Their motion is governed strongly by the fact of their disappearance, regardless of the direction of the surround stereomotion.

Nevertheless, the Object Disappearance Effect occurs only in the three-dimensional context, and is not seen when the dots are viewed in the two-dimensional context of the stimulus on the computer screen. This contextual effect of depth determined by the categorization process is consistent with the Gestaltists' view that *distance (i.e. depth) is a primary and independent perceptual dimension governing the other, dependent perceptual dimensions*. Perceived characteristics of objects are affected by their perceived location in three-dimensional space. The temporally directed, causal sequence of perception implied by perceptual 'equations' thus provides a very strong argument for the existence of indirect perception (Rock, 1997). The primary role of depth may be explained on the base of a 3D-sketch representation or general depth-map, as proposed by Tyler et al (1995) and supported by our previous results (Likova & Tyler, in press).

One might speculate that in the case of a *dynamic* scene the visual system uses abstract 'schemas' of *spatiotemporal* semantic groups as a basis for the event classification. These classification schemas may be involved in the process of applying heuristic assumptions. The Object Disappearance Effect could be considered as a demonstration of the visual system interpreting events by applying specific heuristics. For the case of apparent motion, for example, it has been found (Nakayama et al., 1989; Shimojo et al., 1989) that *depth*-ordering cues provide a context for the correct interpretation of ambiguous motion information, by allowing classification of image features as either intrinsic or extrinsic to the moving surface.

The neural basis for the contextual interactions between *motion* and *depth* was explored by Duncan et al. (2000) in a variation of the classical barber-pole illusion. Their perceptual effect implies a sophisticated interaction between depth and motion information. A plausible site for this interaction is MT (the middle temporal area of the cortex), because information about direction of motion and binocular disparity is known to converge within this area (Maunsell & Van Essen, 1983; Bradley et al., 1995; Bradley & Andersen, 1998; DeAngelis et al., 1998; DeAngelis & Newsome, 1999). These authors examined the sensitivity of MT neurons to the contextual manipulations demonstrated to influence perception. It was found a subset of MT neurons exhibited directional selectivity consistent with perceived surface motion rather than with the motion of the image features present in their receptive field. “These cells properly distinguished between the motions of intrinsic and extrinsic terminators on the basis of *depth*-ordering information” (Duncan et al. 2000). Moreover, they found that depth-ordering information outside the classical receptive fields could be used to resolve the ambiguous motion information found within.

In summary, the illusory depth motion that we report seems to be a manifestation of the general strategy of the visual system to use statistical properties of the environment as heuristics evoking the best explanation of visual events such as a disappearance of an object. This novel dynamic stereophenomenon reveals the operation of cognitive factors in associating depth motion and event perception with object-level processing.

5. ACKNOWLEDGMENT

Supported by NIH grant EY 7890

6. REFERENCES

1. Bradley, D.C., Qian, N. & Andersen, R.A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature* **373**, 609-611.
2. Bradley, D. C., Chang, G. C. & Andersen, R. A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature* **392**, 714-717.
3. DeAngelis, D. L., Gross, L. J., Huston, M. A., Wolff, W. F., Fleming, D. M., Comiskey, E. J. & Sylvester, S. M. 1998. Landscape Modeling for Everglades Ecosystem Restoration. *Ecosystems* **1**: 64-75.
4. DeAngelis, G.C. & Newsome, W.T. (1999) Organization of disparity- selective neurons in macaque area MT. *J. Neurosci.* **19**, 1398-415.
5. Duncan R.O., Albright, T.D. & Stoner G.R. (2000) Occlusion and the interpretation of visual motion: Perceptual and neuronal effects of context. *J. Neurosci*. **20**:5885-97.
6. Duncker, K. (1929/1937). Induced motion. In W.E. Ellis (Ed.), *A Sourcebook of Gestalt Psychology*. London: Routledge and Kegan Paul.
7. Gogel, W. C. (1977) Depth-adjacency and cue effectiveness. *Journal of Experimental Psychology,* **92**(2), 176-181.
8. He, Z. J. & Nakayama, K. (1994). Apparent motion determined by surface layout not by disparity or by 3-dimensional distance. *Nature,* **367**, 173-175.
9. Likova L.T. & Tyler C W. Peak localization of sparsely sampled luminance patterns is based on interpolated 3D object representation (*Vision Research, in press*).
10. Maunsell, J.H.R. & Van Essen, D.C. (1983) Functional properties of neurons in the middle temporal area of the macaque monkey. II Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, **49** 1148-1167.
11. Michotte, A., (1963) *The Perception of Causality,* Methuen and Co. Ltd., London.
12. Minev, K. & Likova, L. T. (1999) Autostereograms as a research tool in stereoscopic vision: Interactions between some cues in perception of motion-in-depth. *Perception* 28, Supplement, 135b.

13. Nakayama, K., He. Z. J. & Shimojo, S. (1995) Visual surface representation: A critical link between lower-level and higher-level vision. In Koslyn, S.M. & Osherson D.N. (Eds.) *Visual Cognition.* MIT Press: Cambridge Ma.
14. Palmer, S. E. (1995) Gestalt psychology redux. In Baumgartner, P. & Payr, S. (Eds.), *Speaking minds*. Princeton, NJ: Princeton University Press, 156-176.
15. Palmer, S. E., (1999). *Vision Science – photons to phenomenology*. Cambridge, MA: MIT Press.
16. Ramachandran V.S. & Anstis, S. M. (1983) Displacement thresholds for coherent motion in random-dot patterns. *Vision Res.* **23**, 1719-1724.
17. Ramachandran V. & Inada V. (1985) Spatial phase and frequency in motion capture of random dot patterns. *Spatial Vision* **1**, 1969-1975.
18. Ramachandran V.S. & Cavanagh P. (1987) Motion capture anisotropy. *Vision Research* **27**, 97-106.
19. Shimojo, S., Silverman G. H. & Nakayama K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Res.* **29**:619-26.
20. Torralba, A. & Oliva. (2002). Global depth perception from familiar structure. Abstract from *Visual Science Society Meeting*. In press.
21. Tyler, C. W. & Kontsevich, L. L. (1995). Mechanisms of stereoscopic processing: stereoattention and surface perception in depth reconstruction. *Perception,* **24**, 127-53.