The Illusion of Perceived Metric 3D Structure

Mats Lind
Dept. of Information Science
Uppsala University
Uppsala, Sweden
mats.lind@dis.uu.se

Geoffrey Bingham
Dept. of Psychology
Indiana University
Bloomington, Indiana
gbingham@indiana.edu

Camilla Forsell
Dept. of Information Science
Uppsala University
Uppsala, Sweden

Abstract

A large body of results on the characteristics of human spatial vision suggests that space perception is distorted. Recent studies indicate that the geometry of visual space is best understood as Affine. If this is the case, it has far reaching implications on how 3D visualizations can be successfully employed. For instance, all attempts to build visualizations systems where users are expected to discover relations based on Euclidean distances or shapes will be ineffective. In that visualizations can, and sometimes do, employ all possible types of depth information and that the results from vision research usually concentrates on one or two such types, three experiments were performed under near optimal viewing conditions. The aim of the experiments was twofold: To test whether the earlier findings generalize to optimal viewing conditions and to get a sense of the size of the error under such conditions. The results show that the findings do generalize and that the errors are large. The implications of these results for successful visualizations are discussed.

1. Introduction

When visualizing multi-dimensional information, the human perceptual system sets a definite limit on the number of dimensions in a data set that can meaningfully be shown. One obvious option is to explore the inherent 3D nature of human perception and use a 3D spatial layout for three important dimensions in the data. This is a standard approach used in numerous visualization toolkits and programs. Research on human spatial vision, however, suggests that the general usefulness of a 3D visual representation may be limited. For instance, space perception researchers have sought the geometry which describes visual space, that is, the geometry that describes the properties preserved over changes in perspective on the space surrounding an observer [1], [2], [3], [11]. The general results from these studies show that space perception is distorted. Most recently, the distortion results have led researchers to suggest that the geometry of visual space is Affine [10], [8]. In Affine geometry, distances along the same direction can be related to one another, but distances along different directions cannot be related. The suggestion is that distance along the depth direction (that is, extending directly away from the eye) cannot be related systematically to distance in a direction perpendicular to the depth direction (that is, in the 'frontoparallel' plane). For instance, if one were looking straight at a wall, the distance to the wall could not be related to a distance extended along the wall itself. The results from the extensive literature of psychophysical studies support the Affine hypothesis: They consistently show distortions in space perception (see [10] for a review of these results.). This might seem puzzling: Our general impression as we look around us is that we have a good sense of the 3D Euclidean structure of the surroundings. One interpretation is that the tasks we are evolved to deal with simply do not require us to make judgements of 3D Euclidean structure. Reaching for something and grasping it, for instance, could be solved by on-line control, something that does not require information about Euclidean structure. All that is needed to start the reaching sensibly is a rough estimate of the distance to the object of interest. Also, non- intuitive results have appeared in other areas of perception research as, for instance, in the case of change-blindness [6].

From a visualization perspective, the lack of information about 3D Euclidean structure means that we can expect users to be unable to correctly compare components of shapes extending in the simulated depth direction to components of shapes defined in the frontoparallel plane. And, therefore, we can expect them to quite often overlook relations in the data that actually exist and are being presented. One remaining issue, and the aim of this study, is to discover exactly how bad Euclidean 3D perception really is under optimal conditions, that is, what the smallest differences in 3D shape are that users are actually able to see in a well formed 3D visualization. The data described above were produced in experiments aimed at testing specific hypotheses about perception. Therefore, the possible sources of 3D shape information, such as shading, texture gradients, ego-motion and binocular stereo, were controlled or varied but never simultaneously present. In the three experiments reported here we used near optimal conditions: Real, structured objects on a visible, structured support surface viewed under normal lighting conditions and the observers were allowed to use normal binocular vision and to move around within certain limits while judging the objects. The first experiment was designed to investigate the size of the errors under such near optimal conditions, the effect on errors of seeing objects under different visual angles and, finally, the effect of different shapes. In the second experiment a wider range of shapes was used and a pure 2D condition was tested to facilitate a direct 2D/3D comparison with respect to errors. The third experiment aimed at testing the possible benefit of including comparison blocks in the presented scene.

2. Experiment 1

The purpose of this experiment was, first, to get an estimate of the size of the errors made when judging 3D shapes under near optimal conditions. Second, we tested the effect of seeing the objects at different distances, and thus under different visual angles. We hypothesized that the errors would increase with the distance to the objects even when the distances differed by relatively small amounts. This would be predicted if the errors were caused by variations in resolution of the optical image.

2.1 Method

Stimuli. Five elliptical cylinders were used, as illustrated in Figure 1. Elliptical cylinders were used because their elliptical shape can be defined in a single measure, the relation between the two axes of the ellipse. The cylinders were hardwood, painted flat black with phosphorescent dots painted on them. They were seen, one at a time, on a tabletop placed at chin height for each subject. This means that the top of each cylinder was seen very slightly from above. Cylinders, having a flat top surface, contain more structure than general smooth objects and, as such, might yield less errors in shape judgments [4]. We used these shapes in an attempt to obtain the best shape perception results that we could. Allowing a slightly visible flat top produces a visible planar contour. The angle between the line of sight, when the top of a cylinder was fixated, and the vertical direction was approximately 86 degrees. The cylinders were all 6.7 cm wide, each with one of five different depths: 3.6 cm, 4.6 cm, 5.6 cm, 6.7 cm and 7.5 cm, yielding depth-to-width (dw-) ratios of 0.54, 0.69, 0.84, 1.0 and 1.12, respectively. Two viewing distances were used: 85 cm and 130 cm. This variation yielded, apart from the obvious distance and retinal size variations, two different angles between the line of sight and the vertical direction when the top of a cylinder was fixated. At 85 cm this angle is approximately 85 degrees and at 130 cm it is about 87 degrees. The two visual angles produced by seeing the objects at the two different distances were 4.5 and 3 degrees of visual angle, respectively.

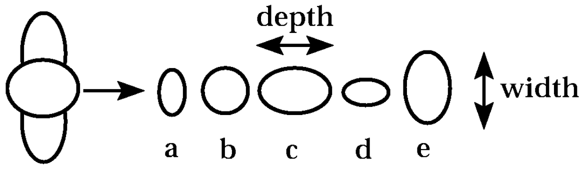

Figure 1. A schematic representation of the objects used in experiment 1 and how their shapes are described.

Apparatus, task and viewing conditions. A computer program was constructed that showed a black ellipse on a white background. The ellipse size and shape could be altered by means of keys on the keyboard. This program was run on a laptop computer with a high resolution TFT screen (1600 by 1200 pixels). The stimulus cylinders were placed on a table adjusted such that its top surface was at chin height for all subjects. The subjects sat some 40 cm away from the table. This gave room for a smaller and lower table to be placed under the table carrying the stimuli. On this smaller table a laptop computer was placed and its screen oriented such that it was approximately perpendicular to each observers line of sight when the observer was looking down at the screen. During each trial the observers were allowed to move when viewing the stimulus cylinder but only with their upper body.

Experimental design. The study was designed as a four factor mixed factorial design with shapes (cylinders), long or short viewing distance and replicates as within-subject factors. The between subject factor was order of presentation for the two viewing distances. Half of the subjects started with the short distance and the other half started with the larger distance. The subjects were randomly assigned to these groups. As mentioned above, five shapes were used. Each subject saw these five shapes in random order three times in each of the far/near conditions. Thus, a total of 30 shapes were judged by each subject (5x3x2).

Observers. Eight observers, three male and five female, took part in the experiment. They were all students at Uppsala University and aged between 21 and 30. When asked, they all reported that they had perfect vision, in some cases with the aid of glasses. They received a small compensation for taking part in the experiment.

Procedure. As the observers entered the room no cylinders were visible. On the table where the stimuli later were presented, a large piece cardboard was standing upright on a stand. The experimenter showed the response apparatus to the observer explained the task and then the experiment commenced. For each trial, the randomly selected cylinder was placed behind the piece of cardboard, the piece of cardboard removed and the observer was asked to produce a picture of the stimulus as seen from above on the computer screen. No time limit was used. When the observer indicated that she or he was satisfied with the ellipse produced on the computer screen, the experimenter again placed the piece of cardboard between the observer and the stimulus.

2.2 Results

Each elliptical shape produced by each observer was described by its dw-ratio. Each shape was shown three times to each observer at each of the two distances. For each distance, observer and shape, the root-mean-square error of the produced shape was computed and compared to the shown shape. These 80 numerical values (5 shapes x 2 distances x 8 observers) were used in an ANOVA and, using a liberal decision criterion of 5%, the F-values were examined. No significant effects were found, that is, for the time being we adopt the view that neither shape (within the 0.54 to 1.12 range) nor viewing distance (between 85 and 130 cm) affected the errors. The mean rms-error across all viewing conditions and observers was 0.138. The results are illustrated in Figure 2.

2.3 Discussion

The mean rms-error indicates a fairly substantial error in the perception of 3D shape even under these favorable conditions. In fact, if we assume that the distributions are normal, our observers need a difference in dw-value of 0.35 to be right 90% of the time, i.e. they would risk to sometimes confuse even the extreme values in our series of shapes. Interestingly, there was no effect of distance, seeing the object at 4.5 or 3 degrees of visual angle did not matter. This indicates that the problem might be one of processing rather than of measuring.

3. Experiment 2

Using the same basic paradigm as in Experiment 1, this experiment aimed at two things: first, to extend the range of shapes to include both thinner and more elongated shapes; second, to compare the shape errors made in the depth direction to the errors made in the frontoparallel direction.

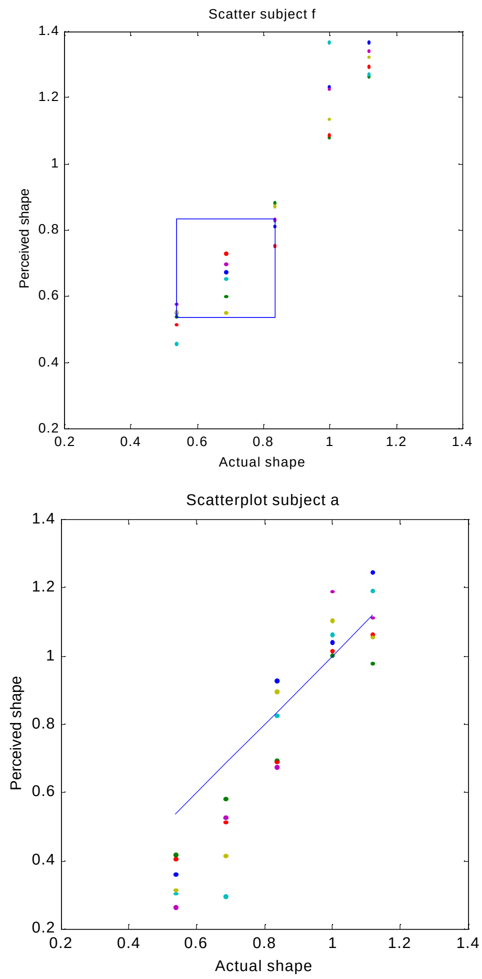

Figure 2. Scatterplots of perceived shape versus actual shape for two representative subjects in experiment 1 giving a view of the sizeable errors made. The solid line illustrates the responses of an ideal observer and is only included for reference.

3.1 Method

Stimuli. Eight elliptical cylinders of the same type as in Experiment 1 were used. They were slightly narrower with widths ranging between 5.4 and 5.8 cm The depths were 2.5 cm, 3.45 cm, 4.5 cm, 5.8 cm, 6.8 cm, 7.65 cm, 9.6 cm and 10.5 cm yielding dw-ratios of 0.46, 0.62, 0.78, 1.0, 1.23, 1.39, 1.66 and 1.81, respectively. Only one viewing distance was used, approximately 85 cm. Apparatus, task and viewing conditions. The same computer program and screen as in Experiment 2 were used. Like in the previous experiments, two tables were used. One larger placed with its top at chin height and one smaller placed in front and below the other table. In the 3D judgement condition, the stimulus cylinders were placed on the top surface of the larger table and the response computer on the lower table. In the 2D condition, the response computer was placed at the front end of the larger table while the elliptic cylinders were placed at the lower table. In the 2D condition, the observers stood up and leaned forward slightly so that they viewed the cylinders directly from the top. The computer screen was then angled such that it was approximately perpendicular to the observer's line of sight.

Experimental design. The study was designed as a three factor within subjects factorial design with shapes (cylinders), replicate and 2D/3D as factors. Each observer saw the eight shapes in random order two times in each of the 2D/3D conditions. Thus, a total of 32 shapes were judged by each subject (8x2x2).

Observers. Ten observers, four male and six female, agreed to take part in the experiment. They were all students at Uppsala University and aged between 21 and 28. When asked, they all reported that they had perfect vision, in some cases with the aid of glasses. However, one subject reported having vision problems as a child and was therefore omitted. For one observer there was equipment failure such that not all judgements were properly recorded by the response apparatus and the data from this observer were discarded. In total, data from eight observers were collected and used. The observers received a small compensation for taking part in the experiment.

Procedure. The procedure was the same as in experiment 1 for the 3D condition. In the added 2D condition, the experimenter placed the piece of cardboard on the smaller table and, to make absolutely sure no preview from an unwanted position could take place, asked the observer to turn around. Behind the cardboard, the cylinder was placed, the observer was then asked to face the table again, to assume the correct position and, finally, the piece of cardboard was removed. The observer after this was asked to produce a picture of the, now visible, top surface of the stimulus cylinder on the response computer. The 3D condition always preceded the 2D condition.

3.2 Results

As in Experiment 1, root-mean-square errors with respect to actual shape were computed for each observer and each viewing condition. These 128 numerical values (8 shapes x 2 viewing conditions x 8 observers) were used in an ANOVA and, using the same decision criterion as before, the F-values were examined. Both the shape factor, the 2D versus 3D condition as well as the interaction were significant (Shape: $F(1,7)=21.0$; $p < 0.005$. 2D/3D: $F(7,49)=3.6$; $p< 0.005$. Interaction: $F(7,49)=3.3$; $p< 0.01$). These effects are illustrated in Figure 3. The effect of shape and the interaction were examined by looking at the slopes of the lines describing actual versus perceived shape. This revealed that the slopes were significantly larger than 1, that is, there was a systematic component in the errors.

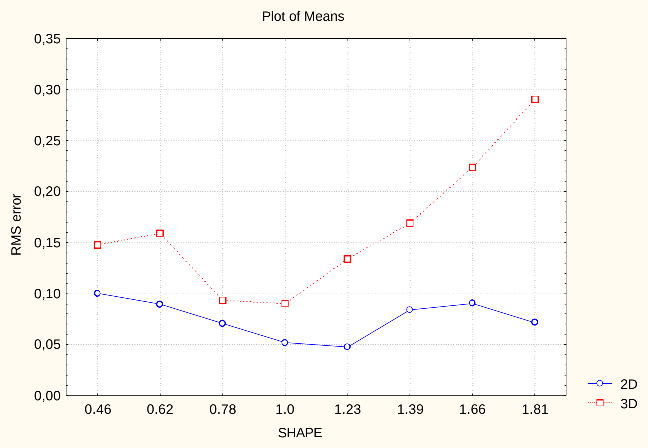

Figure 3. Group mean RMS-error as function of shape for the 2D and 3D conditions in experiment 2.

3.3 Discussion

Judgement of metric shape is more error prone in 3D than in 2D. Furthermore, the errors seem to increase as the shapes become more elongated in depth. Even if this may be partially due to a systematic distortion, and as such possible to remove by preprocessing of 3D scenes, the lowest rms-error values in this experiment are very similar to the ones found in the previous one.

4. Experiment 3

In an effort to find ways of reducing the found errors in metric shape judgements, the idea of using identical comparison blocks was tested. The rationale behind this variation can be found in [9] where the authors propose that metric errors in shape judgements are due to a scaling factor of distances in depth not being recovered by the visual system. If this were the case, a comparison block available in a scene could allow subjects to scale depth in units of the comparison block. If an identical comparison block is then shown in an another scene, the shape of objects in this scene could be compared along the depth dimension to objects in the first scene and metric shape comparisons would be possible.

Having a real comparison block beside the elliptic cylinder to be judged, but a drawing of one on the computer screen next to the response ellipse might induce suspicion in some subjects that there was some kind of manipulation of the comparison figure. To convince the subjects the comparison blocks were identical, real blocks were used, not only for the stimuli, but also for the responses and we let the subjects hold and investigate these comparison blocks before the comparison condition commenced.

4.1 Method

Stimuli. Three elliptical cylinders, a-c, were used. Object b was circular (depth and width both =7cm). Objects a and c were elliptical. By rotating these two objects five different dw-ratios were defined (see figure 1). These depth to width ratios were 0.83, 0.88, 1.0, 1.14, and 1.20. The comparison blocks were cardboard boxes 5.5 cm wide, 5.5 cm deep and 2.5 cm high painted in white, red and black. There were four such comparison blocks, two were always presented beside the stimulus cylinder and two amongst the response cylinders.

Apparatus, task and viewing conditions. As in the previous experiments, the stimuli were presented on a structured tabletop positioned at each subjects chin height. In this experiment the distance form the subjects eyes to the stimulus was approximately 85 cm. The response apparatus was a set of 22 cylindrical objects with shape values ranging from 0.35 to 2.8. The subjects' task was to identify the stimulus object amongst the set of response cylinders when these were seen almost directly from above. To ensure this, the response cylinders were placed on a low table directly to the left of the subjects. The subjects were not allowed to touch the response cylinders, only to indicate which cylinder they saw by pointing. The subjects remained seated during the whole experiment. During each trial they were allowed to move when viewing the stimulus cylinder but only with their upper body.

Experimental design. The study was designed as a four factor mixed factorial design with shapes (cylinders), with or without comparison blocks and replicates as within-subject factors. The between subject factor was order of presentation for the with/without comparison blocks conditions. Half of the subjects started with the condition having comparison blocks and half of them started in the condition without comparison blocks. As mentioned above, five views of three cylinders were used. Each subject saw these five shapes in random order three times in each of the with/without comparison blocks condition. Thus, a total of 30 shapes were judged by each subject (5x3x2).

Subjects. Eight subjects, all male, took part in the experiment. They were students or staff of the Swedish National Defense College aged between 24 and 40. When asked, they all reported that they had perfect vision, in some cases with the aid of glasses. They received no compensation for taking part in the experiment.

Procedure. As the subjects entered the room no cylinders except the response cylinders were visible. On the table were the stimuli later were presented, a large piece of cardboard was standing upright on a stand hiding whatever was behind it. The experimenter showed the response cylinders to the subject, explained the task and then the experiment commenced. For each trial, the randomly selected cylinder was placed behind the piece of cardboard, the piece of cardboard removed and the subjects asked to look at the stimulus and then point at a cylinder in the response set being identical in shape. No time limit was used. When the subject had pointed at a response cylinder, the experimenter again placed the piece of cardboard between the subject and the stimulus and then recorded which cylinder in the response set that was chosen by the subject. After every third trial, the relative position of the cylinders in the response set was changed randomly. When comparison blocks were used, these were placed on each side of the stimulus cylinder with a fixed distance of 10 cm between them and with the stimulus cylinder centered in between them and at randomly selected positions amongst the response cylinders.

4.2 Results

The resolution of the response measure, giving the subjects a finite set of response blocks to choose from, is of course lower than the method used in the previous experiments. In addition, the number of shapes to choose from was not evenly distributed over the range. Therefore a more appropriate response measure for these conditions was chosen, the number of erroneous selections for each subject in each of the two main conditions: With or without comparison blocks. The median number of erroneous selections was 8 out of 15 possible in the condition without comparison blocks. In the condition with the comparison blocks the median number of erroneous selections was 10 out of 15 possible. There are no theoretical reasons to believe that comparison blocks would make the errors larger, but out of curiosity a statistical analysis of the differences were carried out using the Wilcoxon matched-pair test, employing, as before, a decision criterion of 5%. The found difference was not significant. As a complement, the root-mean-square errors were computed. This analysis too showed slightly larger errors in the conditions with the comparison blocks. No statistical analysis was made on these data.

4.3 Discussion

Using physical objects not only for the stimuli but also for the selection of responses seemed a very natural task to subjects in that they just had to point to a cylinder they thought was the same as the one they saw on the tabletop. The results are from that perspective solid. The idea that comparison blocks could serve a purpose in visualizations designed for metric shape judgements receives no support whatsoever.

5. General Discussion

The results from reduced viewing conditions in basic perception research generalizes well to more representative full-cue situations and paint a fairly gloomy picture in terms of what people can perceive of metric 3D structure in visualizations. The human visual system does not seem to be constructed for making metric judgements in 3D space. If we wish to require people to make metric judgments in applications, then we will need to provide differences in depth-to-width ratios of 35% to 50% to allow the differences to be perceived. In practice, this is useless. We have to find other uses for computer generated 3D representations than landscapes where the user is supposed to discover relations based on Euclidean distances or shapes. Ordinal structure in 3D, especially if juxtaposition is present, for instance, is much easier for the visual system to interpret and appreciate and could be used as it is already in many cases. Examples are the ideas behind Data Mountain [5] and 3D tree views [7]. Possibly the perception of metric 3D structure can be improved if additional features such as regular gridlines, yardsticks and the like are added, although the results form experiment 3 cast some doubts on that. This will be further investigated in forthcoming work. However, including formal and numerical components may push viewers into a more sequential and analytic mode of thinking that may be counterproductive in relation to the task at hand.

6. Acknowledgements

This research was sponsored by project AQUA at the Swedish National Defense College in Stockholm, Sweden.

References

[1] Baird, J.C. & Biersdorf, W.R. (1967). Quantitative functions for size and distance judgments. *Perception & Psychophysics* , 2, 161-166.

[2] Gilinsky, A. (1951). Perceived size and distance in visual space. *Psychological Review* , 58 , 460-482.

[3] Gogel, W. C. (1977). The metric of visual space. In W. Epstein (Ed.), *Stability and Constancy in Visual Perception*. New York: Wiley.

[4] Pizlo, Z. (1994). A theory of shape constancy based on perspective invariants. *Vision Research*, 34(12), 1637-58.

[5] Robertson, G., Czewinski, M., Larson, K., Robbins, D., Thiel, D., & van Dantzich, M., (1998) Data Mountain: Using spatial memory for document management. In: *Proceedings of UIST'98*, San Fransisco, CA, November 1998, ACM Press, pp 153-162.

[6] Simons, D. (2000) Current Approaches to Change Blindness, In: Simons, D. (ed) (2000) *Visual Cognition: Special Issue ¾ Change Blindness and Visual Memory*, 7.

[7] Tavanti,M. & Lind M., (2001) 2D vs. 3D, Implications on Spatial Memory, In: Andrews, K., Roth, S. & Wong, P. C. (eds) (2001) *Proceedings of 2001 Information Visualization*, IEEE Computer Society.

[8] Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, 21 (3), 663-678.

[9]Todd, J.T. & Bressan, P. (1990). The perception of 3-imensional affine structure fromminimal apparent motion sequences. *Perception & Psychophysics* , 48, 419-430.

[10] Todd, J. T., Tittle, J. S., & Norman, J. F. (1995). Distortions of three-dimensional space in the perceptual analysis of motion and stereo. *Perception,* 24 , 75-86.

[11] Toye, R.C. (1986). The effect of viewing position on the perceived layout of space. *Perception & Psychophysics* , 40, 85-92.