# Accuracy of Visual Depth Perception in an Open Field

Lauren Beck[1] and Toshiro Kubota[2]
[1]Biology Department, Susquehanna University, Selinsgrove PA 17870
[2]Mathematical Sciences Department, Susquehanna University, Selinsgrove PA 17870

## Abstract

We investigated how accurately humans could estimate distances in an open space. In particular, we were interested in how multiple landmarks are jointly perceived. Two sets of six randomly shaped markers were used to mark locations. One set was laid out on a flat open area where few landmarks were present and all markers were fully visible from a reference point. The other set was laid on a hill-covered area where many landmarks (trees, a sidewalk, a building, and a lamp post) were present and most markers had their lower portions occluded by the hill. Subjects were asked to stand at the reference point and estimate the distance to each of the twelve markers. Data indicated joint estimates tended to be aggregated into a small number of clusters. When the data was separated by subjects' gender, significant differences between male and female depth perception were highlighted.

## Introduction

At the Grand Canyon, one's initial visual impression of the vivid scenery is that it both spectacular and awe-inducing. Unfortunately, disappointment often ensues as snapshots are unable to capture the same degree of grandeur that up-close tourists remember. But why does such a discrepancy between visual experience and photographic capture exist? According to award-winning photographer Robert Caputo of *National Geographic*, "when we look at a landscape, our eyes travel over it and selectively focus on the elements that we find appealing. Our field of vision encompasses a great deal of the scene, but our eyes and brains have the ability to ignore all except the most alluring details. Lenses and sensors or film cannot do this by themselves. They need help," (Caputo, 2007). So does a photograph capture rays of photons more faithfully than our perception which "ignores all except" a few details? Is the Grand Canyon's undeniable grandeur merely a result of our imagination?

Another example of this phenomenon can be seen in the size of a moon. When the full moon rises near the horizon, it appears larger than when it is near the zenith. However, if the full moon is photographed sequentially as it rises from the horizon, it is clear that its size does not change. This is referred to as the moon illusion.

From these two examples, a natural question arises. Do our perception and photography differ vastly in commonplace, everyday settings? Or are the landscape letdowns and full moon illusion special cases? According to J. J. Gibson, the difference lies in the observer. Whereas we actively explore the environment around us, pictures are unable to reproduce this complete three-dimensional view. Rather, they are time-frozen snapshots of the environment (Gibson 1979, Cutting 2003). However, this theory provides no explanation for the nature of these differences, nor does it speculate on their magnitude. If Caputo's view is correct and we construct the Grand Canyon from a small set of landmark points, there may be some evidence within our depth perception as to why these exceptions occur. It is responsible for mentally reconstructing the three-dimensional structures of the canyon, and thus may explain why the photographic images are less spectacular in appearance.

Considering these instances, the general goal of this research is to study how accurate/inaccurate human depth perception is when little depth cues are present. Past research shows that our depth perception is unreliable, but do the inaccuracies occur in a predictable fashion? If so, do the underlying rules predictably contribute to the exaggerated perception of grand landscapes suggested in Caputo's statement? To investigate this problem, we measured the accuracy of humans' depth perception and searched for any patterns evident within the resulting data.

Visual depth perception involves the ability to see three-dimensionally, including both an objects' volume and its spatial distance relative to the viewer. In humans, this ability is especially interesting since our entire three-dimensional perception is based entirely on a two-dimensional image projected onto the eyes' retina. The rods and cones of the retina convert the image's light into electrical signals which are sent to the visual cortex area of the brain's occipital lobe. However, this simplification only explains the mechanisms responsible for 2-D vision, not how depth perception is achieved. One of the most widely-accepted approaches to reconciling this dilemma involves examining perceptual depth cues, which are responsible for transforming a human's surroundings into three-dimensional sight (Mikhailova, 2012).

There are a great many types of such cues, but they can be divided into three broad categories—oculomotor, monocular, and binocular. Oculomotor cues are based on the position of the two eyes and tension within the eye's muscles, and include convergence (inward movement of eyes that occurs when looking at something close) and accommodation (the eye lens' changing of shape) (Cutting and Vishton, 1995). Monocular cues, also called pictorial cues, are still present when vision is limited to only one eye (Turnbull, 2004). Texture gradient, familiarity of relative size and height, perspective convergence, occlusion (nearer object blocking part of a more distant one), shading and shadowing, atmospheric perspective (distant objects have a blue tint and are fuzzier) and motion parallax are all specific examples of pictorial cues. Early artists including DaVinci used these cues to create the illusion of depth in their paintings (Wade, 2012).

The main binocular cue is binocular disparity, the slight differences in the retinal images of the right versus left eye. These result from the different positions of the eyes--the average of 6 cm of distance that separates each eye on an adult human. Stereopsis is the impression of depth that results from this disparity. In 1679, early on in vision research, French mathematician Georges Louis LeClerc diagrammed retinal disparities. But it was not until 1838 that his work was put into practical application with English scientist Charles Wheatstone's invention of the stereoscope. This instrument creates the illusion of depth by overlapping figures with precise horizontal disparities, paving the way for future technological advancements in vision research.

In terms of depth perception physiology, much research has focused on finding the neurons responsible for interpreting binocular disparity. One important discovery was of binocular depth cells, specific neurons located in the striate cortex of the brain. These neurons fire when presented with left and right stimuli with an absolute disparity of about 1 degree (Uka, 2003). Researchers have also located neurons responsible for some pictorial cues, including those of texture gradient. Today, a combined neurological and cue-based explanation for depth perception offers great promise (Tsutsui, 2005).

Though we are not aware of any published research focusing on gender differences in depth perception, there is ample evidence indicating significant differences in the visio-spatial processing of males and females. In particular, males are far superior at tasks involving mentally rotating three-dimensional figures (Mikhailova, 2013). Though the reasons for these differences have not been determined with any degree of certainty, data suggests they are physiological in nature. Hemisphere organization may be responsible, though it's likely only a piece in the puzzle. Males tend to rely heavily on the right hemisphere of the brain when performing spatial tasks. In contrast, females have shown to use both left and right hemispheres to process tasks bilaterally (Rilea, 2008).

Our depth perception experiment was held in a grassy field on the campus of Susquehanna University. Markers were placed across twelve locations, six of which were in an open flat field and six of which were in a hilly area. Subjects were asked to stand at a reference point and estimate the distance from the point to each of the twelve markers. On the open flat field, the ground was fully visible, thus could provide some depth cues. The hill was also surrounded by objects which could potentially supply depth information. The experiment was repeated over two days, and the distributions of the markers were altered. We studied not only the accuracy of estimation, but also the patterns of separations between the depth estimates. Gender differences in depth perception were also considered.

## Materials and Methods

Twenty-six subjects were brought in over a period of two days—July 10$^{th}$ and July 11$^{th}$ 2013. Across the two days, 13 (4 and 9) were male and 13 were female (9 and 4).The subjects were recruited on a volunteer basis and were not compensated. Twenty two out of twenty six volunteers were Susquehanna students working on the campus during the summer. The average age of the twenty-six subjects was 22.0with a median of 20.  They had limited knowledge of what the study was regarding, with vague references to "visual perception" being the sole information given. All subjects had normal or corrected to normal vision. The experiments were held at approximately 5:30 pm on Smith Lawn of the Susquehanna University campus on clear, sunny days.  Smith Lawn is in front of the Cunningham Arts Center and adjacent to the Smith dormitory building.

In order to examine the depth perception accuracy of active observers, twelve depth markers were used. Six markers were inserted into the grass on a flat area of Smith Lawn (Figure 1), and another six were inserted into a hilly area (Figure 2). The markers were made out of poster paper cut into differently sized shapes, attached via duct tape to bamboo poles. The shapes and sizes of the markers were varied to eliminate size as a visual depth cue. Six different colors (red, orange, yellow, green, blue, and purple) were used for identification of the markers. Examples can be seen in Figure 3.

**Figure 1.** The markers on the *hilly* part of Smith lawn day one (top left) and day two (bottom left). View from the reference point is seen to the right. Figures magnified to highlight detail.

**Figure 2.** The markers on the *flat* Smith lawn day one (top left) and day two (bottom left). View from the reference point is seen to the right. Figures magnified to show detail.

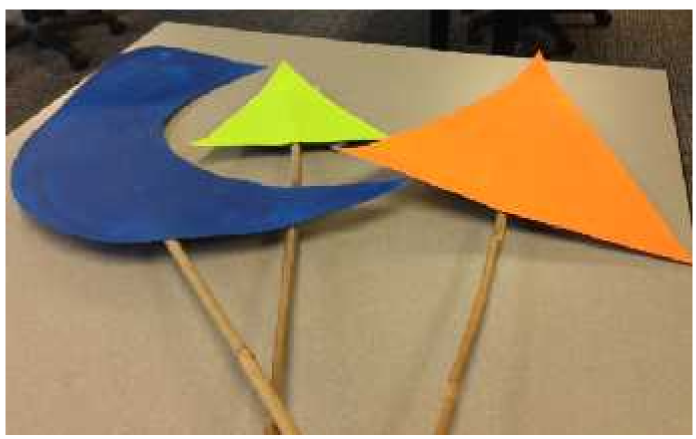

**Figure 3.** General examples of the markers used for experimentation. Shape and color varied.

A pink flag was inserted into the lawn as a reference point. Subjects were asked to stand at the reference point, and using their visual depth perception, estimate the distance to each of the twelve markers. Estimates were handwritten on blank tables prepared and distributed prior to the experiment's commencement. Subjects were permitted to use whichever unit of distance they felt most comfortable with—feet, meters, yards, etc. For the purpose of data analysis and comparison, all of the data were later converted into feet by the experimenters. Subjects were asked not to discuss their estimates with the other participants, and were given the opportunity to ask questions about the procedure. Subjects took approximately ten minutes to complete the entire trial.

Layout of the markers was altered between day one and day two of experimentation. Additionally, since the open area of Smith provided more space, the six markers used there were further away from the reference point, and generally more spread apart than those on the hill. The exact distances of the markers from the reference point are given in Table 1.

**Table 1.** The actual distances the twelve markers were from the reference points across the two days of experimentation. The distances are listed in order from farthest away to closest from the reference point. All units are in feet (ft).

| Hilly Area-Day One | Flat Area- Day One | Hilly Area-Day Two | Flat Area-Day Two |
|---|---|---|---|
| 160 | 300 | 160 | 300 |
| 130 | 250 | 155 | 290 |
| 110 | 210 | 145 | 270 |
| 95 | 180 | 130 | 240 |
| 85 | 160 | 110 | 200 |
| 80 | 150 | 80 | 150 |

To determine these distances, a surveyor's wheel was used. The experimenter stood at the reference point and then rolled the surveyor's wheel out from reference point flag to the appropriate distance. There, the experimenter inserted the marker into the ground. The shaped markers were staggered so as not to occlude one another, thereby eliminating a potential unintended depth cue.  This process was repeated for each marker.  Once the data from the twenty-six subjects was collected, it was manually entered into Microsoft Excel for applicable analysis and graphing.

# Results

## General study

One female subject estimated two of the markers at one yard away (converted to three feet) which seemed improbable. We assumed she mistook the yard/feet conversion factor, and thus removed her responses altogether in our analysis. As a result, responses from 25 subjects (12 females and 13 males) were considered.

In order to analyze the data in a unit-independent manner, they were normalized. The normalization process involved dividing the subject' estimates by the estimate of the marker nearest to the reference point. The normalization was done for each subject and for each area separately. The actual distances to the markers were also normalized by dividing them by the distance to the nearest marker. Figure 4 shows plots of average normalized estimates at each normalized marker location, separated by areas (left) and days of the experiments (right). In all cases, average estimates deviate away from the true value as the marker distance increases.

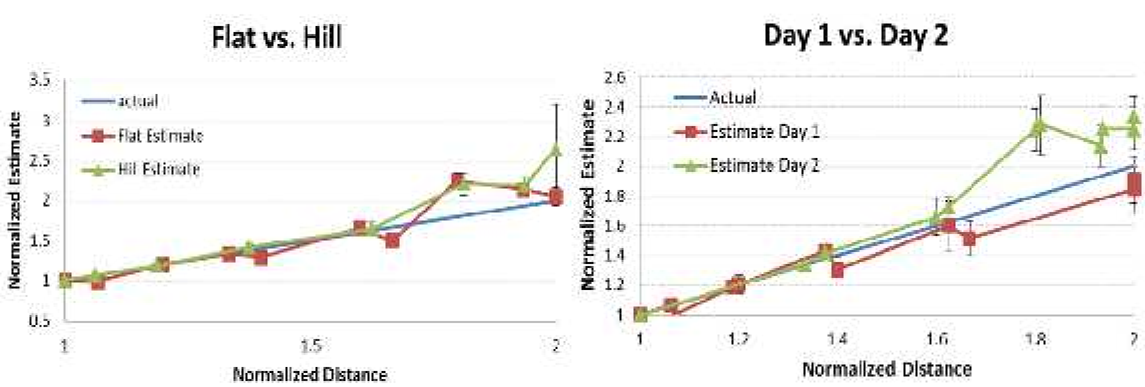


**Figure 4**. Comparison of depth estimates between areas (left) and distributions (right) with standard error bars.

We were interested in examining the surrounding's effects on estimates. As noted previously, the flat area provided few surrounding objects and a clear view to the ground. In contrast, the hilly area provided many surrounding objects but an occluded view of the ground.

Furthermore, due to space constraints, the distances to markers in the flat area were about twice as large as the distances to corresponding markers in the hilly area. We measured the accuracy of depth estimates by the absolute difference between an estimate and its corresponding marker distance after normalization. We called this difference the *absolute error*. To compare the two areas, the means of the absolute errors between the two areas were considered. Table 2 shows sample means and standard deviations of absolute errors in the two areas. There was no statistically significant difference according to a pooled t-test ($p=0.95$).

**Table 2.** Statistics of absolute errors for each area (Flat and Hill).

| | N | Mean | Std |
|---|---|---|---|
| **Flat** | 125 | 0.30 | 0.31 |
| **Hill** | 125 | 0.30 | 0.36 |

Next, we investigated the effect of the distribution of the markers. As shown in Table 1, separations between adjacent markers were all different. On day 1, spacing between the reference point and markers was gradually increased so that the markers closer to the reference point were placed more closely. On day 2, spacing was gradually decreased so that the markers farther from the reference point were placed more closely. Statistics of absolute errors separated by day are shown in Table 3. The mean absolute error in day 2 was larger than that of day 1 ($p=0.083$). We attribute the difference to the fact that markers in Day 1 were closer in average than markers in Day 2; the average distance of the 12 markers in Day 1 was 159 feet while that in Day 2 was 186 feet.

**Table 3.** Statistics absolute errors in each experiment day.

| | N | Mean | Std |
|---|---|---|---|
| **Day 1** | 120 | 0.263 | 0.243 |
| **Day 2** | 130 | 0.337 | 0.401 |

## Gender study

Since males and females have shown notable differences in their visually guided behavior (Mikhailova, 2013), the effect of gender on depth estimation was also examined. Estimates across the days and areas of lawn were pooled, and then separated by the subjects' gender. Both unnormalized and normalized estimates were graphed, as seen in Figure 5.

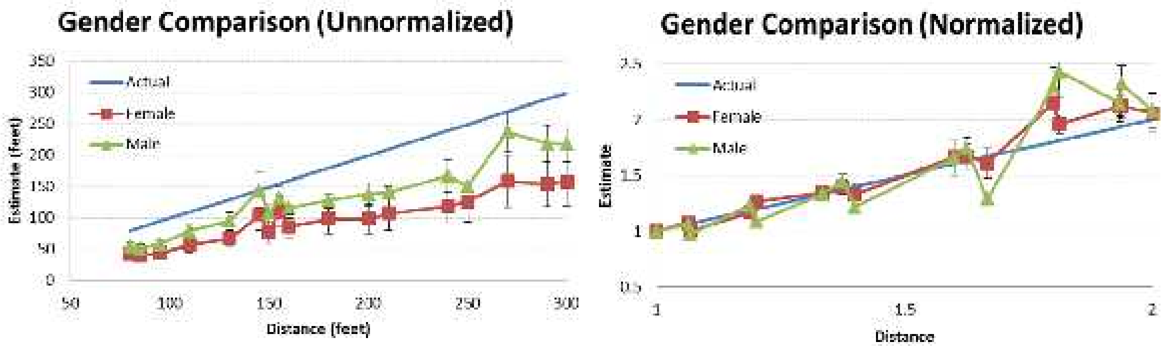


**Figure 5**. Examination on the effect of gender on distance estimations. In terms of the actual metric distance to the markers, both males and females tended to underestimate, but males tended to be more accurate (left). In terms of the normalized distance, females collectively showed slightly more accurate estimates than males. Standard error bars are shown.

From these two plots, striking differences emerge. Collectively, males appear more accurate in estimating distances than females when compared in actual feet (Figure 5, left). However, the trend reverses when comparisons were made using normalized distances (Figure 5, right). We wanted to determine if gender differences persist at the individual level. For this purpose, we calculated average absolute error for each subject. In other words, for each subject, we computed the average of absolute errors over 12 estimates before normalization and 10 data points after normalization (excluding two estimates used for normalization). Figure 6 shows the results in both feet (bar) and normalized unit (line) for each subject, separated by gender. The error bars show the standard error. Table 4 shows the sample means and sample standard deviations of each gender. In feet, the mean of average absolute errors is larger for females than males ($p=0.12$). However, with normalization, the mean is larger for males than females ($p=0.22$).

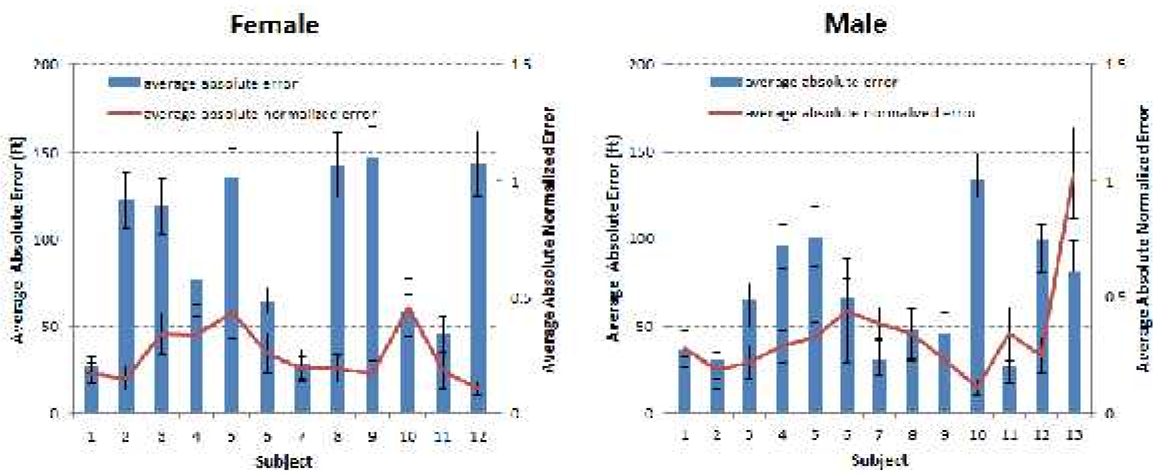


**Figure 6.** Average absolute errors between subjects. Again, data was separated by gender. Standard error bars included.

**Table 4.** Statistics of average absolute estimation errors, separated by gender.

| | | Feet | | Normalized | |
|---|---|---|---|---|---|
| | **n** | **mean** | **std** | **mean** | **Std** |
| **Female** | 12 | 92.45 | 46.92 | 0.25 | 0.11 |
| **Male** | 13 | 66.42 | 33.80 | 0.34 | 0.22 |

## Distribution study

In this subsection, we studied how the estimates were distributed. We were not interested in the numerical accuracy of the depth estimates, but rather how they were spaced out. To determine this, we compared the normalized estimates of pairs of landmarks. The normalization made the data scale invariant, and pairwise differences made the data translation invariant. Scale invariance means that whether an object is visually near to them (larger), or far from them (smaller), the subject can still recognize it. Along the same lines, translation invariance states that an object remains recognizable whether in front of, to the left of, or to the right of, the subject. We considered *every* possible pairwise difference the so results would not be solely due the particular pairwise selection method used. By taking differences of all pairs within the same area (flat or hill), each subject contributed 2x15=30 values.

For each pair of markers, we took the estimates of the markers after normalization, and computed the absolute difference. We called the absolute difference *estimated separation*. The estimated separations were calculated from all subjects and both locations. Then, for each value of the actual depth separation of two markers, the mean of corresponding estimated separations was computed. Figure 7 plots the actual depth separation vs. the mean estimated separation, with error bars showing the standard error. Thus, the figure shows how each depth separation of markers was perceived by the subjects on average. On the y-axis, the average estimated separations were projected, and on the x-axis, the actual separations were projected. We can see that average estimated separations are separated into about five or six clusters, while the actual separations are more dispersed. Indeed, when the k-mean clustering algorithm is applied to the average estimate separations into five clusters, we get clusters shown by red solid ellipses. When the same k-mean clustering is applied to the actual separations, we get clusters shown by green dashed ellipses. Hence, the clusters of the average estimated separations are not due to marker separations but due to some other characteristic of subjects' estimates. In the next section, we discuss what this pattern may indicate.

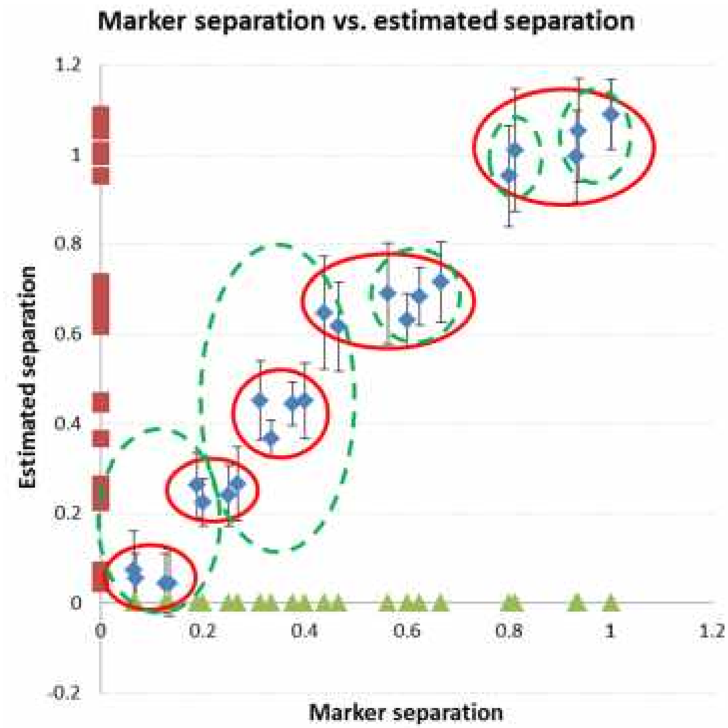


**Figure 7.** A scatter plot of pairwise marker separation and the mean estimated separation for the pair of markers (after normalization). Red ellipses show clusters by k-means using the mean estimate separation. Green ellipses show clusters by k-means using the marker separations. The y-values and x-values of the scatter plot are projected onto the respective axis.

# Discussion and Conclusion

## Gender study

Arguably, vision is of humans' most utilized and important senses, which gives such findings practical implications. Used in simple (walking) and more complex (driving) navigational tasks, one's visual ability can even have an effect on occupational selection. Artists, engineers, and architects, to name a few, must constantly take a step back from their work and look at the overall picture. Additionally, in physically demanding activities such as running, kayaking, and hiking, estimating distance to the "finish line" is crucial for success. Thus, gender differences in depth and distance perception may have even greater societal implications than immediately apparent.

Previous studies have highlighted the differences in male and female visual processing, particularly as it relates to mental rotation tasks (Mikhailova, 2013). Our results show that males were more accurate in estimating metric distances than females. However, after normalization, females were more accurate than males. The data tells us that males are better at pointing to an object and saying, “that chair is 50 feet away”, but females are just as accurate at giving the location of the chair relative to the other objects surrounding it.

But why is this? Males could simply be more familiar with metric distance measurements than females, or the differences could be physiological in nature. Stacy Rilea (2008) studied the sex differences in spatial ability according to a lateralization approach. She found that males process visio-spatial information using their right hemisphere, whereas females tend to complete tasks using a bilateral approach. So, perhaps estimating metric distances requires the right hemisphere more, whereas seeing objects in relation to one another, a “whole picture” so to speak, requires both the left and right hemispheres processing the information symmetrically. Far more complex experiments would need to be conducted to determine for certain which hemisphere was most challenged by our particular experimental design. Milivojec, Johnson, Hamm, and Corballis (2003) proposed that in more difficult spatial tasks, females were just as accurate as males since bilateral processing was required of both genders.

## Distribution study

Figure 7 shows that, collectively, subjects tended to distribute depth separations into a small set of clusters. There are two issues in this interpretation. The first is that we examined the separation of the depth markers relationally, when the data collected were estimates taken from a specified reference point. The second is that we looked at the collective average of estimated separations. Thus, the summarization may not accurately reflect the characteristics of individuals’ estimates. Indeed, there are large variations in the estimated separations. In regards to the first issue, we think that subjects used the separations of landmarks or existing cues to estimate the depth from the reference point. When we look at distant objects, we do not immediately perceive their absolute distances. The gross inaccuracy of estimates in our data supports this claim (Figure 6). Instead, we use separations of objects to incrementally reach the distance estimates from the reference point. From this perspective, estimated separation addresses more directly the characteristics of our depth perception. For the second issue, there are large inter-subject and inter-view variations, but our intra-subject intra-view data do show clustering behavior. Some examples are shown in Figure 8. Thus, the clustering trend is present in each individual and each view.

By further looking at Figure 7, we notice that the width of the clusters (along the x-axis) tends to increase with the actual separation distance. This can be attributed to the Weber effect—the just-noticeable difference of presented stimuli will be proportionally constant with the original size of the stimulus.  Since Weber's law applies to size perception, it is natural to extend its context to depth perception. The law has several implications within our perception of distant objects. When we look at distant landmarks, we try to estimate their separations so that we can reconstruct the scene's structure. We are less interested in the actual distances to these landmarks, since a constant change to the distances only contributes to the translation of the structure. Our binocular vision is not acute enough to resolve these separations precisely. Instead, we resolve them into a few clusters or sparsely distributed bins. Due to the Weber's law, the resolution of the bins decreases with increased separation. When we reconstruct the structure from the landmarks by some type of extrapolation, the result is more coarse and rugged than what it is in reality. This may explain the way we perceive the Grand Canyon and its photographic version so differently.

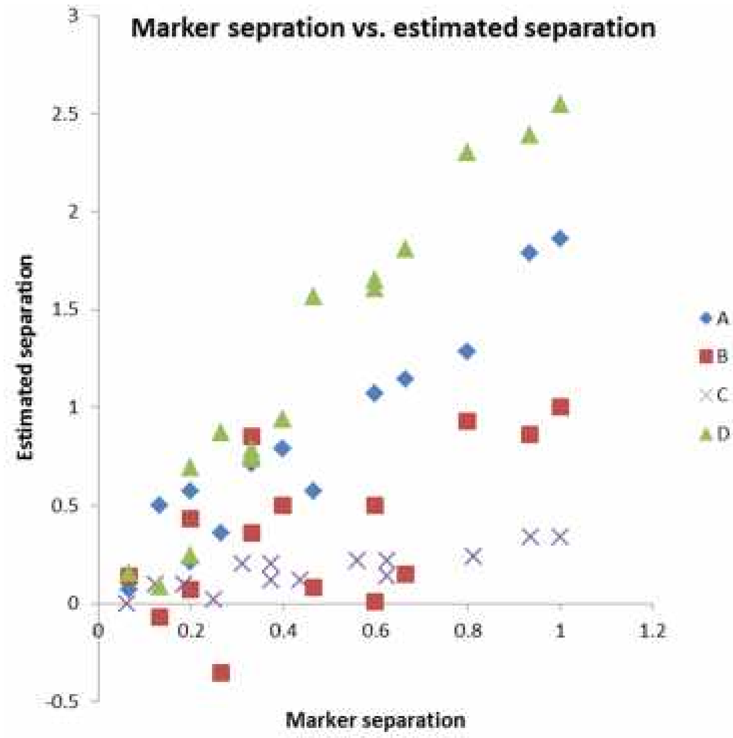


**Figure 8.** Scatter plots of pairwise marker separation and estimated separation for four subjects (A-D) from a single view (Flat or Hill). We see clustering behavior, especially in large marker separations.

## Acknowledgement

This research is supported by the National Science Foundation (CCF-1117439).

## References

Mikhailova, E., Slavutskaya, A., and Gerasimenko, N. 2012. Gender differences in the recognition of spatially transformed figures: behavioral patterns and event- related potentials (ERPSs). *Neuroscience Letters.* 524: 74-78.

Rilea, Stacy. A lateralization of function approach to sex differences in spatial ability: a reexamination. 2008. *Brain and Cognition.* 67: 168-182.

Wade, N. and Ono, H. 2012. Early studies of binocular and stereoscopic vision. *Japanese Psychological Research.* 54: 54-70.

Gogel, Walter C. 1963. Size cues and the adjacency principle. Federal Aviation Agency: Civil Aeromedical Research Institute. CARI report 63-28.

Holway, A. and Boring, E. 1941. Determinants of apparent visual size with distant variant. *The American Journal of Psychology*. 54: 21-37.

Witt, J., Stefanucci, J., and Rien, C. 2007. Seeing beyond the target: environmental context effects distance perception. *Perception*. 36: 1752-1768.

Turnbull, O., Driver, J., and McCarthy, R. 2004. 2D but not 3D: pictorial-depth deficits in a case of visual agnosia. *Cortex*. 40: 723-738.

Julesz, Bela. 1986. Stereoscopic vision. *Vision Research.* 26:1601-1612.

Cutting, James E. 1995. Potency and contextual use of different information about depth. *Perception of Space and Motion.* 69.

Tsutsui, K., Tiara, M., and Sakata H. 2005. Neural mechanisms of three-dimensional vision. *Neuroscience Research.* 51: 221-229.

Uka, T., and DeAngelis, G. 2003. Contribution of middle temporal area to coarse depth discrimination: Comparison of neuronal and psychophysical sensitivity*. Journal of Neuroscience*. 23: 3515-3530.

Caputo, R. 2007. <http://photography.nationalgeographic.com/photography/photo-tips/landscape-photography-tips/>.

Gibson, J.J. 1979. *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Cutting, J. E. 2003. *Reconceiving perceptual space.* H. Hecht, R. Schwartz, & M. Atherton (Eds.) Perceiving pictures: An interdisciplinary approach to pictorial space Cambridge, MA: MIT Press. 215-238.

Li, Zhi, John Phillips, and Frank H. Durgin. 2011. The underestimation of egocentric distance: Evidence from frontal matching tasks. *Attention, Perception, & Psychophysics*. 73: 2205-2217.