Copy-paste in depth

Maarten W.A. Wijntjes[a]

[a]Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands

ABSTRACT

Whereas pictorial space plays an important role in art historic discussions, there is little research on the quantitative structure of pictorial spaces. Recently, a number of methods have been developed, one of which relies on size constancy: two spheres are rendered in the image while the observers adjusts the relative sizes such that they appear to have similar sizes in pictorial space. This method is based on pair-wise comparisons, resulting in n(n-1)/2 trials for n samples. Furthermore, it renders a probe in the image that does not conform to the style of the painting: it mixes computer graphics with a painting.

The method proposed here uses probes that are already in the scene, not violating the paintings' style. An object is copied from the original painting and shown in a different location. The observer can adjust the scaling such that the two objects (one originally in the painting, and the other copy-pasted) appear to have equal sizes in pictorial space. Since the original object serves as a reference, the number of trials increases with n instead of n^2 which is the case of the original method.

We measured the pictorial spaces of two paintings using our method, one Canaletto and one Breughel. We found that observers typically agreed well with respect to each other, coefficients of determination as high as 0.9 were found when the probe was a human, while other probes scored somewhat (but significantly) lower. These initial findings appear very promising for the study of pictorial space.

Keywords: Art and Perception, Depth perception, Pictorial space

1. INTRODUCTION

One of the most prominent themes in western art history is how painters create the illusion of depth. Although many art historians have described the development of pictorial space,[1–3] their conceptions are primarily based on self report. They report what they see in a painting, and although these introspective reports do not cause much controversy, it also does not reveal much about how humans perceive paintings. We will not advocate the introduction of behaviourism in art history, but we do believe that a psychophysical approach to art may lead to new and crucial insights in both art and perception.

In experimental psychology there has been an increasing interest in how to quantify pictorial space. By far the most thoroughly studied topic is that of 3D shape.[4,5] A different geometric aspect that only recently started to receive attention is that of global depth structure.[6] Mainly two methods have been explored. First, a virtual arrow is shown in pictorial space and observers have to point to a target location. The arrow is rendered such that the observer can also point in the 'depth' direction, besides the picture plane dimensions.[7,8] The second method is a relative size task. Here,[9] two disks are shown in the pictorial scene and the observer has to adjust the relative sizes such that both objects appear to have similar sizes in pictorial space. These methods can also be combined and together give quantitative insight in the style of the painting[10] .

Both methods have two potential disadvantages. Firstly, the depth reconstruction formalisms are based on a full data set in which all possible pairs of a certain sampling are measured. Secondly, the probes that are used typically do not conform to the style of the painting. This sometimes results in probes that appear rather awkward in a scene, and may prove to be a distracting factor for the observer. In the study presented here, we explored a new variation of the relative size task in which we use cutouts of the original image as depth probes.

farther author information:
E-mail: m.w.a.wijntjes@tudelft.nl, web: www.maartenwijntjes.nl

The basic idea is that a copy of an object is pasted at a different location in the pictorial scene. The observer has to adjust the size of the pasted object to match that of the original object. The method differers from the original relative size task[9] with respect to the type of probe but also to the number of measurements. In our proposed method, we select n sampling points and only measure once with respect to the standard probe. In contrast, previous methods measured each pair[9] which results in $n(n-1)/2$ trials for n sampling points. The advantage of our method is that it is fast, the disadvantage is that our data may be more noisy.

The purpose of this paper is to explore the copy-paste in depth method. We tested two paintings and perform basic analysis to reveal if results are reliable. Furthermore, we tested whether an original probe (an actual copied polygonal selection of the original painting) gives more reliable results than an 'alien' probe, in our case a rendered sphere. We also explored how to analyse data from different reference probes: in the second painting we performed the experiment three times using three different reference probes. Combination of the data is not straightforward but in the end gives insight in how integrable pictorial space is.

2. METHODS

2.1 Stimuli

Figure 1. Paintings with sampling locations and probes used in the experiment. On the left, 'La piazza San Marco in Venice" (ca. 1723-1724) by Canaletto, copyright at Museo Thyssen. The probes are shown at two times their original scale. Also, the artificial sphere is not present in the original image, but was shown at the indicated location during the experiment. On the right, 'Netherlandish Proverbs" (1559) by Pieter Breughel the Elder, copyright at Gemäldegalerie, Staatlichen Museen zu Berlin - Stiftung Preussischer Kulturbesitz. Photo: Jörg P. Anders. Probes are shown in their original size (relative to the painting).

We performed experiments on two different paintings. The first painting is "La piazza San Marco in Venice" (ca. 1723-1724), one of the many vedutes (city-views) of Canaletto. It depicts the San Marco square containing mainly humans. The scene (at least where we did our measurements) is flat. The paintings seems to be painted well according to the rules of perspective. Note that the San Marco square is actually not rectangular causing the left building block to converge in a different location than the right billing block.

The second painting was "Netherlandish Proverbs" from Pieter Breughel the Elder. It dates from 1559 and depicts more than a hundred Dutch (Nethelandish) proverbs. Because of the subject (many proverbs), the scene is densely packed with a wide variety of objects which makes it a suitable painting to test our copy-paste-in-depth paradigm. Three objects were chosen, a person, a 'ball' and a dog. As opposed to the Canaletto scene, this scene is non-flat, it contains various buildings and the ground itself also has some relief.

The probes were cut out by the author. During preparation, the author manually selected an expansion point from which the probes were scaled. This was the lowest point of the probe (e.g. the foot), making the scaling spatially well defined. The sampling locations were selected by the author. The Canaletto painting had 59 sampling points, the Breughel painting 27.

2.2 Participants

The Canaletto painting was tested on eight observers, one of which was the author of this paper, the rest were naive with respect to the purpose of this study. The Breughel painting was also tested on eight observers. Six observers (including the author) were similar to those for the Canaletto painting, two were new. All observers had normal or corrected to normal vision.

2.3 Procedure

Observers performed the experiment on a computer screen located at approximately 40 cm from their eyes. Observers' heads were not fixed and could freely move. During each trial, a cut out copy of the original probe was shown at a sampling location, initially scaled randomly within 50-150% of the original size. The observers were instructed to adjust the size of the probe such that it appeared to have the same size in pictorial space as the original. The experimenter (the author) verbally ensured whether observers understood that they had to match sizes in pictorial space and not in the (2D) picture plane. Size adjustment was performed using the up and down arrow keys. Each probe was shown in a separate session. Therefore, the Breughel painting was tested in three session, and the Canaletto painting in two session.

2.4 Data analysis

From a linear perspective point of view, the size of the projection (size on screen/canvas) linearly decreases with depth. Thus, if we are completely sure that we are using a proper linear perspective depiction (like a photograph with linear lens) then we can convert the relative sizes in relative distances. However, since we are studying imaginary spaces here (which all paintings are) we do not know the actual perspective. This is not a problem if we only analyse the relative sizes instead of converting them into relative distances.

We focussed our analysis on different topics for the two paintings. For the Canaletto, we analysed whether having an original probe (cut-out from the actual scene) would give different data than an 'alien' probe, a rendered sphere. Next, we analysed whether there were any systematic spatial errors. Since all sampling points are located on the ground, we could fit a plane to the data and quantify whether there were any systematic residuals with respect to the spatial layout of the scene.

For the Breughel we wanted to analyse the effect of using different objects from the scene, the human, ball and dog. We correlated the data between different probes. Since the three reference probes have different locations in the pictorial scene, we analysed how these can be compared and combined in a proper way.

3. RESULTS

3.1 Canaletto

In figure 2A the inter-observer correlations are shown for the two different reference probes. For each observer pair (28 in total) we computed the coefficient of determination (R^2). As can be seen, the data for the person probe show a higher level of agreement between observers than the sphere probe. Statistical significant was confirmed by a t-test ($t(27) = 6.62$, $p < 0.05$).

The San Marco square can be modelled as a plane, slanted in depth. We predict a strong correlation of relative sizes with respect to the height in the picture plane. No correlation in the horizontal direction is expected. In figure 2B the average relative sizes are plotted against their vertical positions in the picture plane. As may be expected, we find a clear relationship with relative sizes increasing for positions lower in the picture plane. Next, we wanted to know whether relative sizes varied in the horizontal direction. To do so, we took the residuals of the model fit shown in figure 2B, i.e. we subtracted the trend in the vertical direction from the data. To our surprise, we found a clear trend in the horizontal direction for the person probe, with a coefficient of determination of 0.62. This trend was practically absent for the sphere probe ($R^2 = 0.08$).

To understand the origin of this unexpected horizontal trend we analysed the original Canaletto painting. The scene contains many human subjects. If the sizes of these humans show the same trend, the observer bias may be explained by that. To quantify the human sizes, the author drew lines over all human subjects which he could clearly see (figure 3A). The length of these lines were digitally converted into relative sizes by normalising

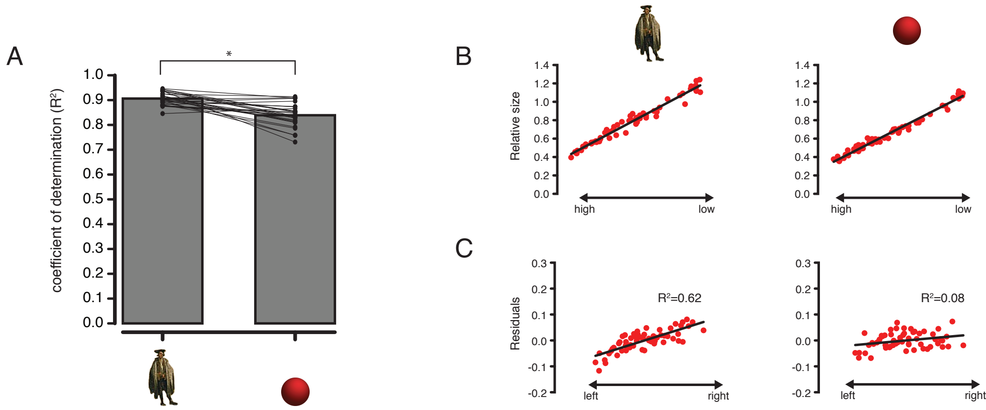

Figure 2. A: Pairwise correlations (R^2) per probe. B: Size correlation with the vertical direction. C: Size correlation with the horizontal direction.

them with respect to the length of the human probe used in the experiment. The resulting data was analysed in the same way as the human observer data. As can be seen in figure 3B, the same vertical trend is found, denoting the diminishing of size with respect to height in the picture plane. These data is somewhat more noisy than the observer data (compare with figure 2B) but this is expected since the various humans may have been depicted with individual heights, as in reality. Importantly, the horizontal trend we found is similar to the observer data. The coefficient of determination is lower, but the slope is higher than the observer data. Mind that we cannot directly correlate the observer data with human size data because the sampling is different: there is are no point-to-point relations.

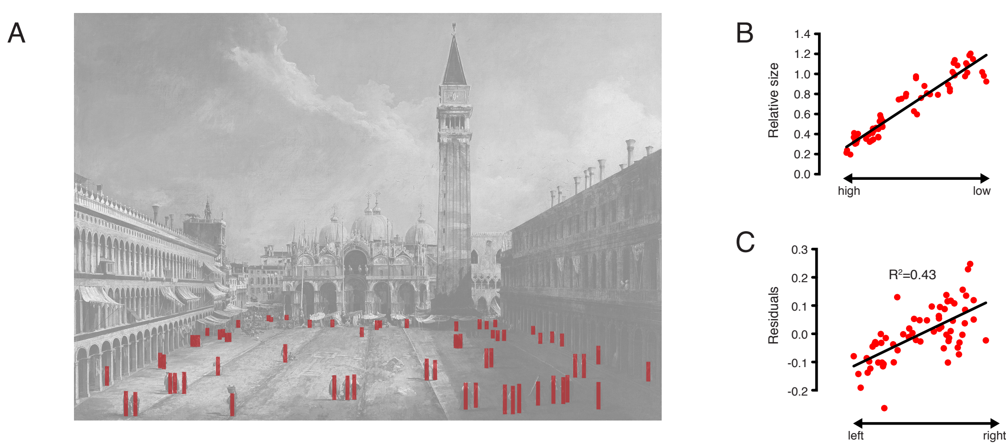

Figure 3. A: Sizes of all humans measured my the author. B: Size correlation with the vertical direction. C: Size correlation with the horizontal direction.

3.2 Breughel

Similar to the Canaletto analysis, we quantified the inter-observer similarity per probe by correlating all observer pairs. The results are presented in figure 4A. A one-way ANOVA revealed a significant main effect of probe ($F(2, 81) = 26.1392$, $p < 0.05$) and Bonferroni corrected pairwise comparisons (at $p < 0.05$) indicted that the human probe showed higher agreement between observers than the other two probes.

Next, we wanted to directly compare the data of different probes. A plot of the average pairwise relative sizes is shown in figure 2. The grey dashed line denotes $y = x$ and the black line is a linear fit to the data. As can be clearly seen, the slopes between the dog probe and the other two is substantially different. At first sight, a different slope may appear to indicate a different depth gain, but that is misleading. To understand these figures one should first understand that any fit should pass through the origin, since this is basically 'infinity', it denotes an infinitely small relative size. The actual meaning of the slope is not depth gain but the 'distance' (in terms of relative size) between two reference positions. For example, the position where the ball has a relative size of 1, the person has a relative size larger than 1 which means that if the person probe would be located at the position of the ball it should increase its size (see figure 4B). All three slopes are in agreement with each other. The first slope indicates that the person is farther than the ball. The second and third slope both indicate that the dog is farther than both the ball and person, but slightly more so with respect to the ball, which is in agreement with the first slope.

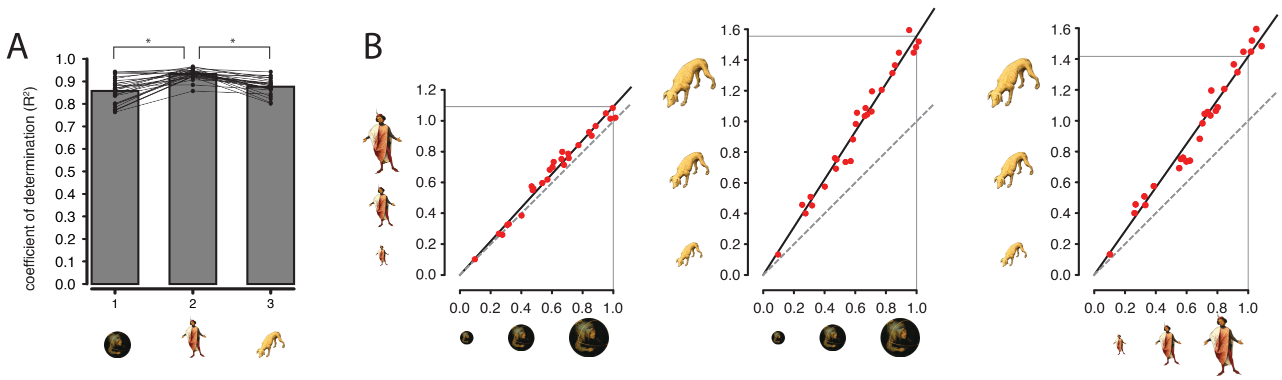

Figure 4. A: Pairwise correlations (R^2) per probe. B: Regressions of relative sizes between the three different probes.

To understand the comparison of different probe data in more detail, we analysed individual data. For each observer we fitted linear functions $y = a_{ij}x$ where a_{ij} denotes the slope between probe i and j. This slope should be indicative of the distance between probe i and j in terms of relative size. We wanted to understand whether these slopes are consistent with respect to each other. As shown in figure 5A, the relative size difference between the dog and ball can be calculated in two ways, directly via a_{13} of via a_{23} multiplied by a_{12}. As can be seen in figure 5B, we can replace the three different probes by three similar probes with well defined relative sizes. To understand whether these triangles were consistent for individual observers we plotted a_{13} against $a_{23}a_{12}$ in figure 5C. It can be clearly seen that all values agree well. The results means that depth in terms of relative sizes is 'integrable', $a_{12}a_{23}a_{31} = 1$ or more general $a_{n,1}\prod_{i=1}^{i=n-1} a_{i,i+1} = 1$.

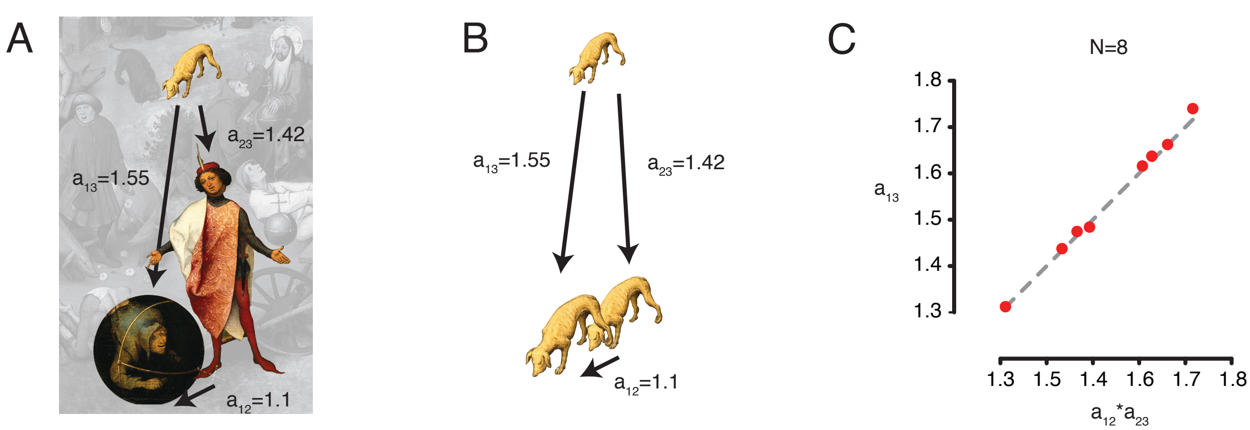

Figure 5. A: The three probes in their original position. The arrows indicate the distances in terms of relative size, which are the slopes shown in figure 4. B: The dog probe is scaled according to the slopes. C: Individual data of the slopes.

4. DISCUSSION

The main results show that there is good agreement among observers. Correlations between observers were rather high, between an average of $R^2 = 0.84$ for the rendered sphere in the Canaletto up to $R^2 = 0.93$ for the person probe in the Breughel. We initially thought that the significant difference between the sphere probe and the person probe ($R^2 = 0.91$) in the Canaletto painting was due to the sphere being 'alien', i.e. not conforming to the style in the painting. However, we found the same difference in the Breughel painting where the sphere was an original item of the scene. Therefore, the difference in the Canaletto is likely due to the difference in object shape or category and may not depend on whether the probe is 'original' or 'alien'. An experiment where the object category is kept the same is probably necessary to quantify the value of original probes as opposed to alien probes.

Interestingly, we find that human probes score a higher similarity among observer than other probes. This effect may have at least two potential causes. Firstly, the human may have a prototypical size that is simply easier to adjust than a ball or dog. Secondly, the pictorial spaces we used in this research are filled with other humans. One could say that human sizes produce the geometry of pictorial space. A straight forward question would be whether size gradients give a global impression of depth on which the observers base their settings, or that observers match human sizes locally to nearby humans. If the latter is true, we would probably find a higher level of noise since figure 3B shows that there is more variability on the actual humans sizes in the painting than is shown in the observer data (figure 2B). On the other hand, the apparent horizontal size gradient we found in the original Canaletto only resulted in a bias for the person probe and not the sphere probe. Therefore, we hypothesise that relative size settings are both based on local and global features.

We showed how data from multiple probes at different reference locations can be combined. For the three probes we found that the data is very consistent. In general, we can quantify consistency by checking whether

$$a_{n,1} \prod_{i=1}^{i=n-1} a_{i,i+1} = 1 \tag{1}$$

Taking into account that $a_{ij} = 1/a_{ji}$ this relationship can be rewritten in $\prod_{i=1}^{i=n-1} a_{i,i+1} = a_{1,n}$ which is what is shown in fire 5C. It should be noted than the multiplication ($\prod$) becomes a summation ($\sum$) when we consider logarithmic relative sizes. Then, (1) reads

$$\log(a_{n,1}) + \sum_{i=1}^{i=n-1} \log(a_{i,i+1}) = 0 \tag{2}$$

This equation is more intuitive since it tells us that the integration along a closed loop of the logarithm of relative sizes should vanish. In other words, the pictorial depth is integrable. Analysis using logarithmic relative sizes or depth has previously also proven useful in research on pictorial space.[6–9]

That the pictorial space of Breughel is integrable does not necessarily generalise to any other painting. Integrability may fail depending on the style of the pictorial space. Therefore, it can be used as a diagnostic tool to quantify various pictorial spaces. Furthermore it should also be noted that, although observers show good agreement with each other, the magnitude of overall depth is different for different observers, This can be clearly seen in figure 5C: Although different paths (a_{13} and $a_{12}a_{23}$) result in similar relative sizes for each observer, there are large difference between observers, ranging from 1.3 to 1.8. This is in line with previous findings on pictorial space.

This research is exploratory and of a rather methodological nature. The results we provide show that it is relatively easy to measure the depth structure of paintings. From an art historic perspective, this is really only a start towards the understanding of depth perception in paintings. The part of our results that is of direct interest to art historians may be the horizontally skewed depth gradient in the Canaletto painting, that was both present in the original painting and in our observer data. What is evidently clear is that there is a large cast shadow at the right side of the San Marco square. Whether Canaletto was simply biased to painting larger figures on a dark background, or whether he had intention to do so is food for speculation. The persons in the shadow seem in general to be of a more aristocratic origin. There also seems to be a higher average of males in the shadow. Details such as the inverted cast shadow directions of the persons in the big shadow seem to indicate that Canaletto thought carefully about the composition, which may support the claim that Canaletto intentionally applied the horizontal size gradient.

The method presented here can also be applied to photos. There is an especially suitable platform for this, which is LabelMe.[11] This open annotation tool is used for various databases that contain huge collections of photos, all annotated by humans. The main purpose of most of these databases is to develop and test computer vision algorithms that learn from human annotations. But a coinciding advantage is that for each annotated object an outline polygon is available. Some occluded objects are less useful but there are many non-occluled objects available, simply due to the massive amount of pictures. The polygon data can be used to cut out an object and paste it in a different location in the scene. In this way, many pictures can be measured without too much of preparatory work by the experimenters.

ACKNOWLEDGMENTS

This research was supported by the Dutch Foundation for Scientific Research (NWO).

REFERENCES

1. White, J., [*The birth and rebirth of pictorial space*], Belknap Press (1987).
2. Willats, J., [*Art and representation: New principles in the analysis of pictures*], Princeton University Press (1997).
3. Dunning, W. V., [*Changing images of pictorial space: a history of spatial illusion in painting*], Syracuse University Press (1991).
4. Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., and Todd, J. T., "Ambiguity and the 'mental eye' in pictorial relief," *Perception* **30**(4), 431–448 (2001).
5. Todd, J. T., "The visual perception of 3d shape," *Trends in Cognitive Sciences* **8**(3), 115–121 (2004).
6. van Doorn, A. J., Wagemans, J., de Ridder, H., and Koenderink, J. J., "Space perception in pictures," *Proceedings of SPIE - The International Society for Optical Engineering* **7865**(786519) (2011).
7. Wijntjes, M. W. A. and Pont, S. C., "Pointing in pictorial space: Quantifying the perceived relative depth structure in mono and stereo images of natural scenes," *Transactions on Applied Perception* **7**(4) (2010).
8. Wagemans, J., Van Doorn, A. J., and Koenderink, J. J., "Measuring 3d point configurations in pictorial space," *i-Perception* **2**(77-111) (2011).
9. Wagemans, J., Van Doorn, A. J., and Koenderink, J. J., "Pictorial depth probed through relative sizes," *i-Perception* **2**, 992–1013 (2011).
10. van Doorn, A. J., Koenderink, J. J., Leyssen, M. H. R., and Wagemans, J., "Interaction of depth probes and style of depiction," *i-Perception* **3**(8), 528–540 (2012).
11. Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T., "Labelme: a database and web-based tool for image annotation," *International journal of computer vision* **77**(1), 157–173 (2008).