Box spaces in pictorial space: linear perspective versus templates

Huib de Ridder*, Sylvia C. Pont

Perceptual Intelligence Lab, Faculty of Industrial Design Engineering,
Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands

ABSTRACT

In the past decades perceptual (or perceived) *image quality* has been one of the most important criteria for evaluating digitally processed image and video content. With the growing popularity of new media like stereoscopic displays there is a tendency to replace *image quality* with *viewing experience* as the ultimate criterion. Adopting such a high-level psychological criterion calls for a rethinking of the premises underlying human judgment. One premise is that perception is about accurately *reconstructing* the physical world in front of you ("inverse optics"). That is, human vision is striving for *veridicality*. The present study investigated one of its consequences, namely, that linear perspective will always yield the correct description of the perceived 3D geometry in 2D images. To this end, human observers adjusted the frontal view of a wireframe box on a television screen so as to look equally deep and wide (i.e. to look like a cube) or twice as deep as wide. In a number of stimulus configurations, the results showed huge deviations from veridicality suggesting that the inverse optics model fails. Instead, the results seem to be more in line with a model of "vision as optical interface".

Keywords: Linear perspective, veridicality, interface theory of perception, foreshortening, viewing distance, object size

1. INTRODUCTION

In the past decades perceptual (or perceived) *image quality* has been one of the most important criteria for evaluating digitally processed image and video content. This resulted in a large variety of standardized methodologies for quantifying human quality judgments. With the growing popularity of new media like stereoscopic displays there is a tendency to replace *image quality* with *viewing experience* as the ultimate criterion[1,2]. Adopting such a high-level psychological criterion calls for a rethinking of the premises underlying human judgment and quality judgment in particular.

One of the most intriguing but often overlooked premises concerns the (usually implicitly adopted) underlying theory of human perception. Currently, the dominant theory of perception states that human perception is about accurately *reconstructing* the physical world in front of you ("inverse optics")[3]. That is, human vision is striving for *veridicality*. Such model is in line with the signal-processing approach towards image quality where it is generally taken for granted that there always will be a reference (for example, in image coding the original image) and that the impairment of image quality due to coding artifacts is related to the distance between the original and coded image[4,5]. Recently, however, the model of vision as inverse optics striving for veridicality has been challenged by the so-called interface theory of human perception[6,7,8,9]. This theory states that perception is not about accurately reconstructing the physical world but about *constructing* the properties and categories of an organism's perceptual world. On evolutionary grounds[6,9], one can argue that these perceptual structures are not intended to accurately match the physical world but, instead, are fast, intention-driven explorations of the meaningless physical world in preparation of "optically guided potential behavior" [7,8], thus striving for utility and efficiency, not veridicality. The efficacy of such mind-to-world approach to perception increases by making use of prototypes or templates like memory colors[10] or canonical visual size[11]. In a similar vein, this approach justifies the existence of reference-free techniques as a viable way of evaluating image quality[12,13].

*h.deridder@tudelft.nl; phone + 31 15 2781815; fax: +31 15 278 7179

Human Vision and Electronic Imaging XVII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Huib de Ridder, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8291, 829117 • © 2012 SPIE-IS&T
CCC code: 0277-786X/12/$18 • doi: 10.1117/12.916464

Recently, Pont et al. [14] devised an experiment in which predictions from both theories could be directly confronted with each other. Their study concerned the depiction of 3D spaces in 2D images. According to the inverse optics model of visual perception such depiction should follow the rules of linear perspective in order to produce a faithful reproduction of the original scene. The experimental results, however, yielded huge deviations from model predictions based on linear perspective suggesting that the observers didn't go for veridicality. Instead, the results were much closer to the predictions on the basis of the interface model of perception supporting the growing experimental evidence for a model of "vision as optical user interface" [7,8].

The objective of the present study is to replicate and extend the experiment by Pont et al. To this end, we investigated a variety of stimulus configurations as well as the impact of changing the instructions to the participants. The results can be found in section 3. As an introduction to our experiments, a brief summary of the study by Pont et al. is presented in section 2. Main conclusions can be found in section 4.

2. DEPTH IN BOX SPACES

As already stated above, the study by Pont et al. [14] was about the depiction of 3D spaces in 2D images and in particular whether linear perspective will always yield the correct description of the perceived 3D geometry in 2D images. In general, this is investigated by taking a static image and vary the viewing conditions, assuming that if the eye is not in the center of projection belonging to that image, the image will look distorted[15]. In the study by Pont et al. the experimental procedure was exactly the opposite: per trial the viewing condition was fixed while the image was interactively manipulated by the observer until the perceived configuration in the image was considered to be in line with the instructions to the observer. The stimulus was a transparent wireframe picture of a frontoparallel box space against a white background (Fig. 1, left-hand side) shown on a 37" wide-screen (aspect ratio 16/9) LCD television set to a resolution of 1280 by 1024 pixels and a vertical refresh rate of 75 Hz. The (invisible) horizon divided the wireframe in two equal parts with the primary vanishing point in the center of the picture. The participants were instructed to adjust the wireframe picture such that it looked like the picture of a cube, i.e. equally deep as wide. This was accomplished by changing the ratio of the sizes of the backside of the box space (i.e. side B of the inner square) and the front (i.e. side A of the outer square) while keeping the outer square constant. According to the rules of one-point linear perspective, the resulting ratio B/A (or foreshortening *f*) will depend in case of a cube on viewing distance L and object size A as follows:

$$f = B/A = L/(L + A) = 1/(1 + A/L). \quad (1)$$

The relations between viewing distance, object size and foreshortening are illustrated in figure 1, right-hand panel. At a given viewing distance, the veridical or true foreshortening will systematically decrease as a function of object size. For a given object size, on the other hand, true foreshortening will increase as a function of viewing distance.

Seven participants, highly trained in psychophysical experiments, performed the experiment at five different viewing distances (30, 60, 120, 240 and 480 cm) with an eye-patch in front of their non-dominant eye. At each distance, they adjusted interactively the foreshortening for five object sizes (2.6, 5.2, 10.5, 20.9 and 41.9 cm). These experimental conditions resulted in true foreshortenings ranging from 0.42 to 0.99. The dashed line in the middle panel of figure 2 denotes the best linear fit to the experimental data averaged across the participants. Its shallow slope indicates that the responded foreshortening only slightly increased as a function of true foreshortening and in fact suggests that the participants preferred a constant value as predicted by the interface theory of vision. This is supported by the fit to the regression line by a weighted sum of true foreshortening f_{true} and constant foreshortening $f_{template}$, or

$$f_{responded} = w.f_{true} + (1\text{-}w).f_{template} \ , \quad (2)$$

weight w being equal to 0.28 and $f_{template}$ being equal to 0.56. In figure 2, the upper left-hand and the lower right-hand wireframe boxes represent the setting for the constant foreshortening $f_{template}$ while the lower left-hand and upper right-hand wireframe boxes represent the corresponding veridical foreshortenings at low and high values of f_{true} respectively. Observe that the wireframe boxes with the constant foreshortening $f_{template}$ look like cubes whereas the corresponding veridical settings appear deformed. That is, at low values of foreshortening the wireframe box looks like an elongated corridor whereas at high values it looks very flattened, lacking a sense of depth. These observations support the suggestion that human observers don’t go for veridicality but rather prefer a template representing a standard view of a cube.

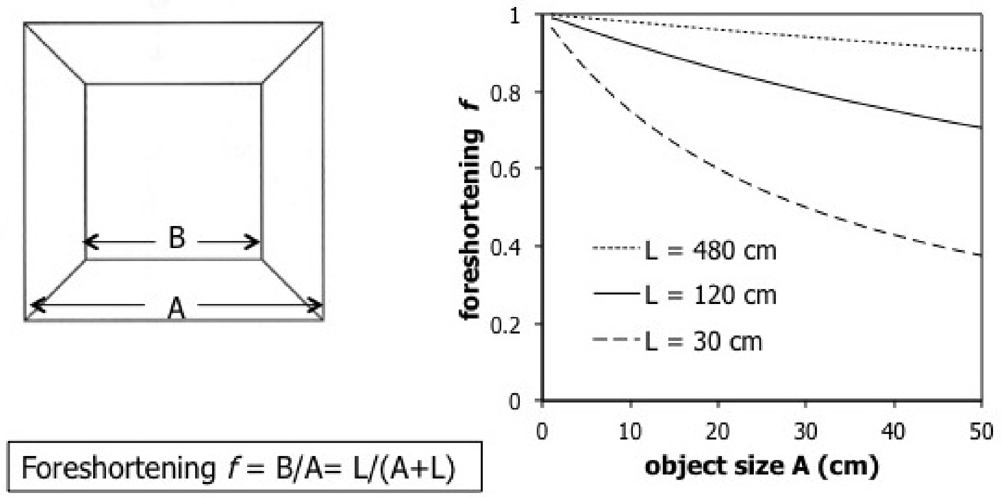

Figure 1: Foreshortening of a wireframe cube (i.e. a box space at aspect ratio one) as a function of object size (A) and viewing distance (L) as predicted by the rules of one-point linear perspective.

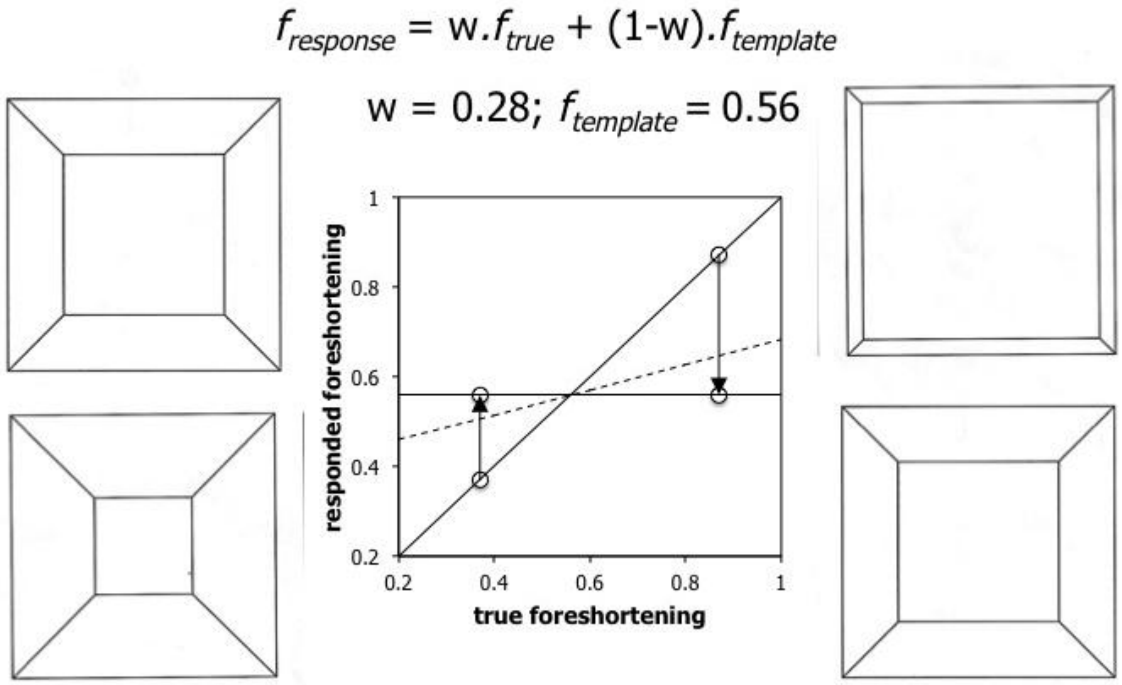

Figure 2: Schematic representation of the experimental results from the study by Pont et al. [14]. The dashed line in the middle panel denotes the best linear fit to the experimental data. The upper left-hand and the lower right-hand wireframe boxes represent the setting for the constant foreshortening $f_{template}$ while the lower left-hand and upper right-hand wireframe boxes represent the corresponding veridical foreshortenings at low and high values of f_{true} respectively.

3. EXPERIMENTS

In order to extend the study by Pont et al. [14], we introduced, in separate sessions, one of the following additions to their experimental setting: (1) a context was added making the horizon and vanishing point explicitly visible, (2) the wireframe box was rotated over 45 degrees transforming it into a diamond-shaped box space. Furthermore, naïve participants performed the experiments. Finally, in one condition the participants were instructed to adjust the foreshortening of the wireframe box space such that it looked either equally deep and wide (i.e. look like a cube, aspect ratio of one) or twice as deep as wide (i.e. aspect ratio of two).

3.1 Method

3.1.1 Participants

All participants were students from the faculty of Industrial Design Engineering, Delft University of Technology, who followed the first-year course "Research & Design" in 2010 and 2011. They had no experience with psychophysical experiments. All had normal or correct-to-normal vision. Some participants viewed the images with two eyes. A meta-analysis showed that this had no impact on the experimental results. All participants mentioned that they experienced a sense of depth; in particular for the large object at short viewing distance it was felt as if they were inside the box space.

3.1.2 Procedure

The experimental set-up was similar to the one in the study by Pont et al. The viewing conditions were confined to three: 30, 120 and 480 cm. In addition, the object sizes were limited to four: 5, 10, 20 and 40 cm. In total, four types of experiments were performed. In the first one, 27 students measured the impact of viewing distance and object size on the foreshortening, thus replicating the study by Pont et al. but now with naïve observers. In the second experiment, ten students evaluated two situations, one with and one without contextual information. In the latter condition, the transparent wireframe box appeared to be placed on a tiled surface with the orthogonal lines converging to the vanishing point on the horizon. Figure 4 contains a screenshot of this configuration. In the third experiment, six students assessed the impact of rotating the wireframe box space by 45 degrees to the original setting. Finally, in the last experiment ten students determined the appropriate foreshortening in two conditions: the original one where the wireframe box should look like a cube (aspect ratio of perceived width and depth being 1:1) and one with aspect ratio of perceived width and depth being 1:2.

3.1.3 Data analysis

The results of the last three experiments were analyzed by means of multiple linear regression using one dummy variable (X = {0,1}) to quantify the impact of the additional variable (context, rotation, aspect ratio) or

$$f_{responded} = a + b.f_{true} + c.X + d.X.\ f_{true}$$

$$X = 0 \quad => \quad f_{responded} = a + b.f_{true}$$

$$X = 1 \quad => \quad f_{responded} = (a + c) + (b + d).f_{true} \quad , \qquad (3)$$

where X = 0 stands for the original condition, i.e. similar to the one in Pont et al., and X = 1 for the new condition.

3.2 Results

Figure 3 presents the results of the first experiment. The relation between the responded foreshortening and the true foreshortening could be fitted by a linear regression line with a slope of 0.21 and an offset of 0.40 ($r^2 = 0.56$; $p < 0.001$). This implies that weight w equals 0.21, which is close to the value determined by Pont et al. (0.28). By means of eq. (3), $f_{template}$ could be established to be 0.50, comparable to the value in Pont et al. (0.56). Apparently, the lack of experience in performing psychophysical experiments had no impact on the results.

Figure 4 presents the results of the second experiment. They could be fitted by eq. (3) ($r^2 = 0.80$; $p < 0.001$) with a = 0.42 ± 0.02 ($p < 0.001$), b = 0.22 ± 0.02 ($p < 0.001$), c = -0.03 ± 0.02 (p = 0.25, NS) and d = 0.06 ± 0.02 (p = 0.11, NS). Apparently, the addition of the context had no effect on the responded foreshortenings. By means of eq. (2), weight w was estimated to be 0.24 with $f_{template}$ being equal to 0.53.

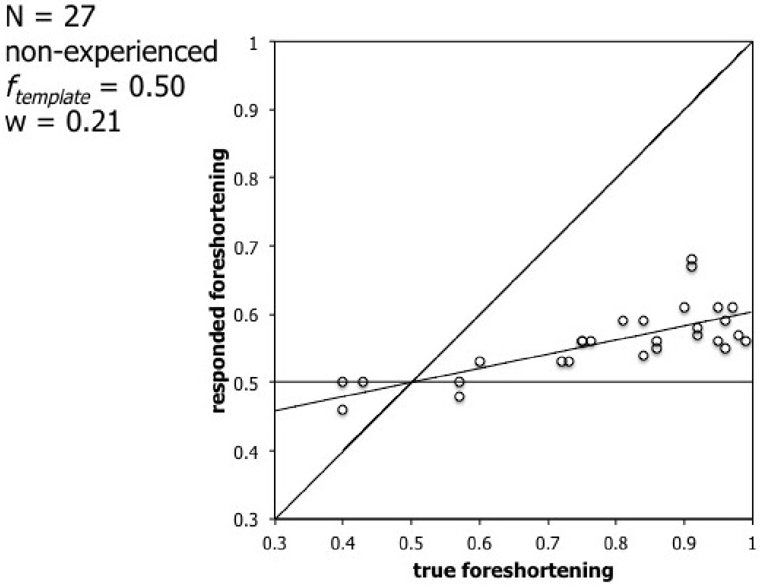

Figure 3: Responded foreshortening as a function of true foreshortening.

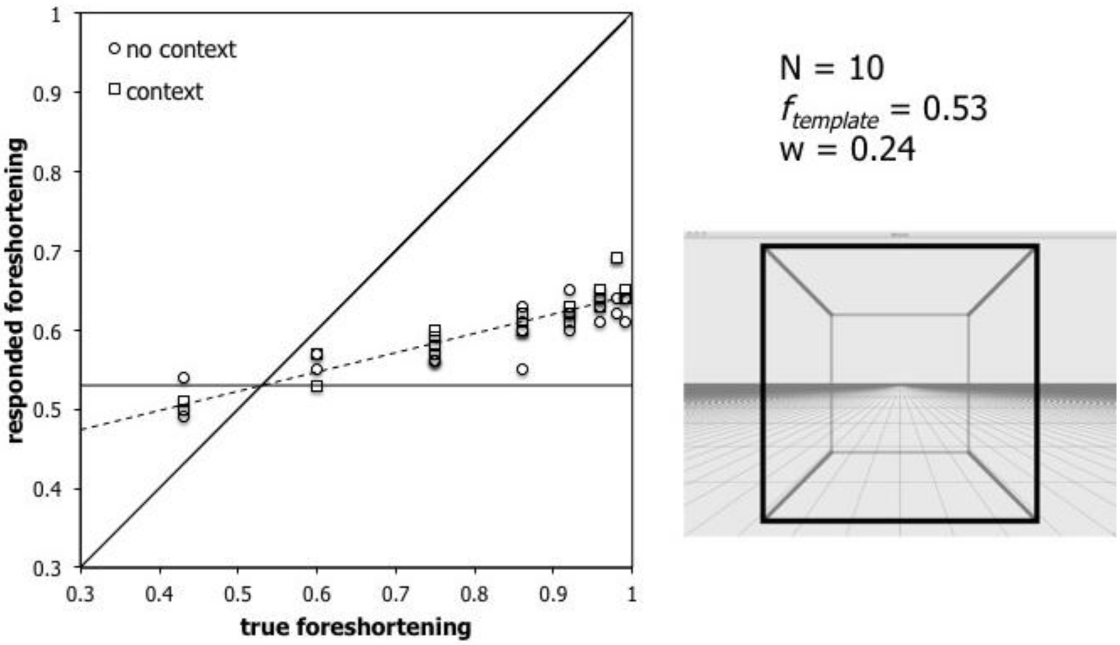

Figure 4: Responded foreshortening as a function of true foreshortening for two contextual situations.

Figure 5 presents the results of the third experiment. They could not be fitted by eq. (3) (r^2 = 0.04; p = 0.15) suggesting that the responded foreshortening is independent of the true foreshortening. But if eq. (3) is still applied to the data, then the resulting parameters are: a = 0.51 ± 0.02 (p < 0.001), b = 0.05 ± 0.02 (p = 0.03), c = 0.04 ± 0.03 (p = 0.16, NS) and d = -0.06 ± 0.03 (p = 0.10, NS). Apparently, the variation in orientation had no effect on the responded foreshortenings. By means of eq. (2), weight w was estimated to be 0.03 with $f_{template}$ being equal to 0.55.

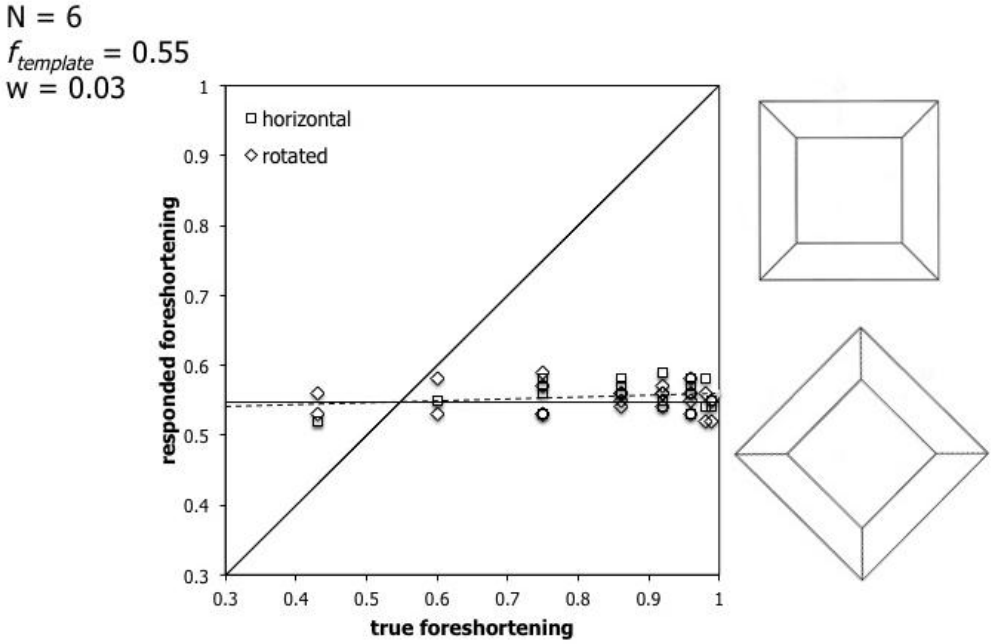

Figure 5: Responded foreshortening as a function of true foreshortening for two orientations of the box space.

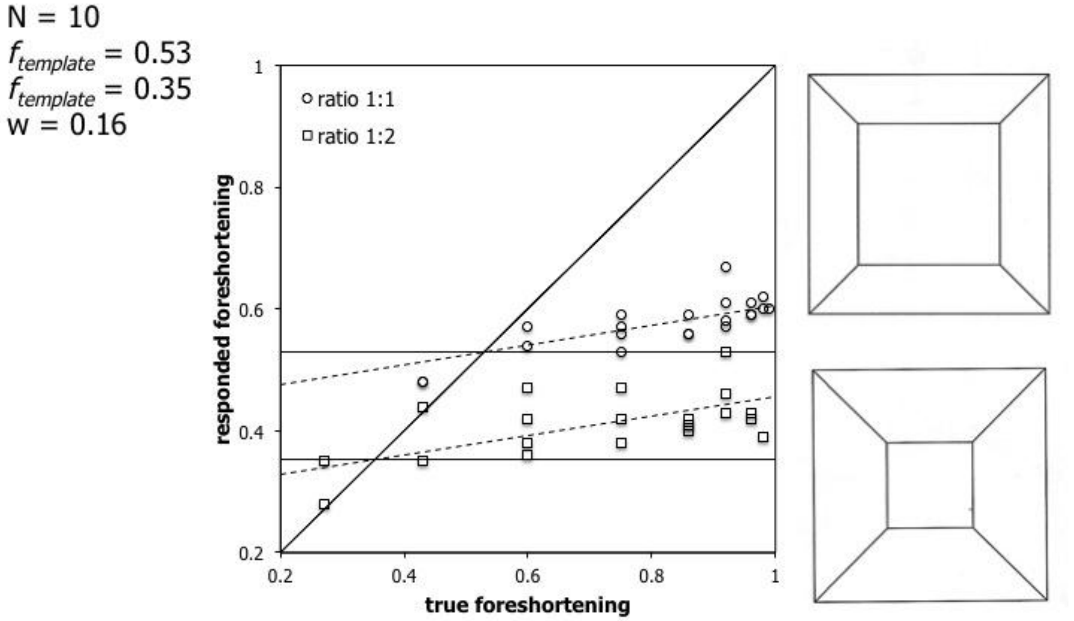

Figure 6: Responded foreshortening as a function of true foreshortening for two instructions about the aspect ratio.

Figure 6 presents the results of the fourth experiment. They could be fitted by eq. (3) ($r^2 = 0.87$; $p < 0.001$) with a = 0.41 ± 0.04 ($p < 0.001$), b = 0.20 ± 0.05 ($p < 0.001$), c = -0.10 ± 0.05 (p = 0.05) and d = -0.06 ± 0.06 (p = 0.24, NS). Note that now the dummy variable has a significant weight, implying that the regression lines for the two conditions are parallel but with a different offset. By again fitting eq. (3) to the results but this time with a reduced set of parameters (i.e. parameter d is left out of the equation), the new weights are: a = 0.44 ± 0.03 ($p < 0.001$), b = 0.16 ± 0.03 ($p < 0.001$), c = -0.15 ± 0.01 ($p < 0.001$). By means of eq. (2), weight w was estimated to be 0.16 with $f_{template}$ being equal to 0.53 and 0.35 for aspect ratios of one and two, respectively. On the right-hand side of figure 6, the upper and lower wireframe box spaces correspond with these values of $f_{template}$. Assuming the aspect ratio for the cube to be one and using the two estimated values for $f_{template}$, the perceived aspect ratio for the other box space was calculated to be equal to 2.07. This value is close to the instructed value of two, suggesting that the aspect ratio is preserved in the present setting.

4. CONCLUSIONS

The results of the present study confirm and extend the conclusions drawn from the previous study by Pont et al[14]. The main conclusions about the correct description of the perceived 3D geometry in 2D images are the following:

- There exist large deviations from veridicality. For the box spaces, the true foreshortenings due to the size-distance effect contributed not much to the responded foreshortenings with weights for true foreshortening (f_{true}) varying between 0.03 and 0.28.
- There is a preference for templates or standard views instead of taking perspective into account. For the box spaces with aspect ratio of one (i.e. cubes), the value of $f_{template}$ varied between 0.50 and 0.56 while its contribution to the responded foreshortenings was large with weights varying between 0.72 and 0.97.
- Aspect ratio of width and depth of a box space was preserved in the current setting.
- Results support the “vision as optical interface” concept.

In short, these conclusions indicate that human observers maintain a notion of how things should look in a “canonical view” [11] and that, if the rendering deviates from this canonical view, they start to complain, in disregard of such parameters as viewing distance and picture size. Painters know this from experience already for a long time and apply corrections to perfect linear perspective in order to make paintings look right[8,15]. For example, in the painting *School of Athens* Raphael drew off-center spheres as circles rather than as ellipses, even though this is the proper shape for their perspective image, in order to avoid the spheres looking distorted[15]. The present study supports this approach in a quantitative way.

REFERENCES

[1] Seuntiëns, P.J.H., Heynderickx, I. and IJsselsteijn, W.A., “Capturing the added value of 3-D-TV: Viewing experience and naturalness of stereoscopic images”, *Journal of Imaging Science and Technology*, 52, 1-5 (2008).

[2] Lambooij, M., IJsselsteijn, W., Bouwhuis, D. and Heynderickx, I., “Evaluation of stereoscopic stills: Beyond 2-D quality”, *IEEE Transactions on Broadcasting*, 57, 432-444 (2010).

[3] Palmer, S.E., *Vision Science: Photons to Phenomenology*, MIT Press, Cambridge, Mass, USA (1999).

[4] Ridder, H. de, “Minkowski-metrics as a combination rule for digital-image-coding impairments”. In: *Human Vision, Visual Processing, and Digital Display III* (B. E. Rogowitz, J.P. Allebach, S.A. Klein, eds.), SPIE Proc. vol. 1666, 16-26 (1992).

[5] Ridder, H. de, “Current issues and new techniques in visual quality assessment”. In: *ICIP-96*, vol.1, 869-872 (1996).

[6] Hoffman, D.D., “The Interface Theory of Perception: Natural selection drives true perception to swift extinction”. In: *Object categorization: computer and human vision perspectives* (S. Dickinson, M. Tarr, A. Leonardis, B. Schiele, eds.), 148-165, Cambridge University Press, Cambridge, UK (2009).

[7] Koenderink, J.J., “Vision and Information”. In: *Perception beyond Inference. The information content of visual processes* (L. Albertazzi, G.J. van Tonder, D. Vishnawath, eds.), 27-57, MIT Press, Bradford Book, Cambridge, Mass, USA (2011).

[8] Koenderink, J.J., “Vision as a user interface”. In: *Human Vision and Electronic Imaging XVI* (B.E. Rogowitz, T.N. Pappas, eds.), SPIE Proc. Vol. 7865, 786504-1-13 (2011).

[9] Mark, J.T., Marion, B.B. and Hoffman, D.D., “Natural selection and veridical perceptions”, *Journal of Theoretical Biology*, 266, 504-515 (2010).

[10] Yendrikhovskij, S.N., Blommaert, F.J.J. and Ridder, H. de, “Representation of memory prototype for an object color”, *COLOR research and application*, 24, 393-410 (1999).

[11] Konkle, T. and Oliva, A., “Canonical visual size for real-world objects”, *Journal of Experimental Psychology: Human Perception and Performance*, 37, 23-37 (2011).

[12] Ridder, H. de, “Psychophysical evaluation of image quality: from judgment to impression”. In: *Human Vision and Electronic Imaging III* (B.E. Rogowitz, T.N. Pappas, eds.), SPIE Proc. Vol. 3299, 282-291 (1998).

[13] Ridder, H. de, “Cognitive issues in image quality assessment”, *Journal of Electronic Imaging*, 10,47-55 (2001).

[14] Pont, S., Nefs, H., Doorn, A. van, Wijntjes, M., Pas, S. te, Ridder, H. de and Koenderink, J., “Depth in box spaces”, *Seeing and Perceiving* (in press).

[15] Todorovic, D., “The effect of the observer vantage point on perceived distortions in linear perspective images”, *Attention, Perception & Psychophysics*, 71, 183-193 (2009).