Peer-reviewed version available at *Vision* **2018**, *2*, 26; doi:10.3390/vision2030026

Article

Multiple pictures of a perspective scene reveal the principles of picture perception

Casper J. Erkelens

Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht
e-mail: c.j.erkelens@uu.nl

Abstract: A picture is a powerful and convenient medium for inducing the illusion that one perceives a real three-dimensional scene. The relative invariance of picture perception across viewing positions has aroused the interest of painters, photographers and visual scientists. Many studies have been devoted to perceptual invariance when pictures are viewed from oblique directions. Invariance across viewing distances has received less attention. This study presents a computational analysis of pictures of perspective scenes taken from different distances between camera and physical objects. Distances and directions of pictorial objects were computed as function of viewing distance to the picture and compared with distances and directions of the physical objects as function of camera position. The computations show that pictorial distance and direction are determined by angular size of the depicted objects. Pictorial distance and direction are independent of camera position, focal length of the lens, and picture size. Ratios of pictorial distances, directions and sizes are constant as function of viewing distance. The constant ratios are proposed as the reason for invariance of picture perception over a range of viewing distances. Reanalysis of distance judgments obtained from the literature shows that perspective space, previously proposed as the model for visual space, is also a good model for pictorial space. The geometry of pictorial space contradicts some conceptions about picture perception.

Keywords: picture perception; pictorial distance; angular size

1. Introduction

Pictures are images on flat surfaces, in which human subjects can see objects at a distance (relative to the viewer) and in depth (i.e. relative to other objects). Defined in this way, pictures are both physical objects (i.e. canvas, paper or screen) and planar representations of retinal images. A convenient aspect of pictures is their viewpoint-independent utility. That is, a viewer need not be directly in front of a picture at the point from which it was taken to enjoy it, to understand it, to admire it, or simply to look at it and make sense of what is seen (Busey, Brady & Cutting, 1990). Picture perception has been studied during oblique viewing. Many authors concluded that viewers compensate for incorrect viewpoints. They advocated theories of picture perception relying on mental operations that rectify the Euclidean geometry of the original scene (Goldstein, 1987, 1988; Halloran, 1993; Perkins, 1973; Pirenne, 1970; Rosinski & Farber, 1980; Rosinski, Mulholland, Degelman & Farber, 1980; Wallach & Marshall, 1986; Yang & Kubovy, 1999; Vishwanath et al., 2005). Busey et al. (1990) claimed that one can look at moderately slanted pictures without perceptual interference because the distortions in the image are sub-threshold, or within the bounds of acceptability. Other authors did not find evidence for view-point-compensation (Koenderink et al., 2004; Rogers & Gyani, 2010; Todorovic, 2008). Generally, picture perception and real-world perception have been conceived as different. Gibson (1979), Sedgwick (2003), Costall (1990), Hagen (1986), Hochberg (1962, 1978), Kennedy (1974), Kubovy (1986), Rogers (1995), Willats (1997) emphasized differences between perception of the world and pictures of it. Koenderink and colleagues proposed that pictorial space should not be thought of as "three-dimensional," but rather as "two-plus-one-dimensional," the single dimension being "depth" (Koenderink et al., 2011).

Pictorial space was described as a fiber space, with the visual field as base space, and the depth dimension as fibers (Koenderink & van Doorn, 2012). The description specified that base and fibers have fully distinct geometrical structures (the base space approximately Euclidean, the fibers close to affine) and are largely independent of each other. Koenderink and colleagues further argued that familiar size is not a distance cue in picture perception (Koenderink et al., 2008; Wagemans et al., 2011). Wagemans et al. (2011) recognize that the size cue is well understood for the perception of physical objects, i.e. for visual space. Familiar size can act as an effective distance cue because the distance from the eye to an object equals the ratio of its physical size to its angular extent in the visual field. Koenderink and colleagues argue that such simple geometrical relations do not apply to pictorial space, since the eye itself is not in pictorial space, and consequently the notion "distance from the eye" is meaningless (Koenderink et al., 2008; Wagemans et al., 2011). The eye not being an object in pictorial space is a fallacious argument because it creates an irrelevant distinction between pictorial space and physical space. The relevant distinction to be made is between physical space on the one side and perceptual spaces, such as visual space and pictorial space, on the other side. The eye is not an object in visual space, but yet numerous studies showed that distance is a useful concept for judging the remoteness of physical objects (Baird and Wagner, 1991; Bian, Braunstein, & Andersen, 2005; Da Silva, 1985; Feria, Braunstein, & Andersen, 2003; Foley, Ribeiro-Filho, and Da Silva, 2004; Gilinsky (1951); Haber, 1985; He & Ooi, 2000; He, Wu, Ooi, Yarbrough, & Wu, 2004; Madison, Thompson, Kersten, Shirley, & Smits, 2001; Meng & Sedgwick, 2001, 2002; Ni, Braunstein, & Andersen, 2004; Ooi, Wu, & He, 2001, 2006; Philbeck & Loomis, 1997; Sinai, Ooi, & He, 1998; Toye, 1986; Wagner, 1985, 2006; Wiest & Bell, 1985; Wu, Ooi, & He, 2004). Eyes and objects have positions in physical space that are represented in the brain. The physical position that most obviously qualifies for being the reference for distance and direction in pictorial space is the position from which the picture is viewed.

In a number of experimental and computational studies, I investigated the geometry of perspective space and proposed it as a model for visual space (Erkelens, 2013a, 2013b, 2015a, 2015b, 2015c, 2017). Experiments included judgments on physical and depicted objects. Two conclusions relevant for the current study were that 1) familiar shape and size are powerful cues for slant and distance perception and 2) apart from a stronger underestimation of slant and distance, there was no reason to assume a different geometry for pictorial space. To further test the hypothesis that the geometries of pictorial and visual space are similar, this study presents computations made on sets of two pictures of a perspective scene containing familiar objects. The pictures were taken from different distances. The computations enable the comparison between the geometries of pictorial, visual and physical space. Comparison of pictorial space with visual space also comes from data in the literature. Data obtained by Kraft and Green (1989) of the perceived distance of depicted objects as function of their physical distance is fitted to the perspective model of visual space (Erkelens, 2017). The computations of perceived distances and directions in this study are based on the following hypothesis: "When looking at a picture, viewers perceive the distance of a depicted object (the *physical* distal stimulus) as the distance of an imaginary physical object (the *pictorial* distal stimulus) that produces the same retinal image (the *proximal* stimulus)". The hypothesis proved to be successful in describing perceived slant of obliquely viewed grid figures as functions of depicted slant and slant of the picture (Erkelens, 2013a, 2013b).

2. Comparison of pictures taken from different camera positions

Figures 1(a) and 1(b) show two pictures of the same perspective scene taken with a digital SLR camera (Nikon D5100). Size of the camera's APS-C sensor is 15.7 x 23.6 mm so that a focal length (FL) of 36 mm corresponds with a field of view (FoV), defined as the diagonal angle of view of the lens, of 43^{o}. An FL of 72 mm corresponds with a FoV of 22^{o}. Figure 1(a) was taken with the FL 36 mm lens and Figure 1(b) with the FL 72 mm lens. The pictures were taken from two camera positions such that objects nearby the camera, e.g. the traffic signs, were projected to similar locations in the pictures. By printing both pictures at the same size, the traffic signs have the same size in both pictures. The traffic signs appear also at the same distance when you look into the pictures. Far objects have

Peer-reviewed version available at *Vision* **2018**, *2*, 26; doi:10.3390/vision2030026

different sizes in Figures 1(a) and 1(b). The house at the left side is smaller in Figure 1(a) than in Figure 1(b). The house in Figure 1(a) is seen at a larger distance than the house in Figure 1(b). The difference in distance seems to support the general opinion that lenses of different focal lengths make a scene look compressed or expanded in depth (Cooper et al., 2012). Short lenses expand depth whereas long lenses compress depth. A pertinent question is why this is the case. Figures 1(c) and 1(d) show that depth compression is not equivalent to distance compression. Figure 1(b) has been resized to Figure 1(c) such that the houses in Figures 1(a) and 1(c) are equal of size. Figure 1(c) is a factor of 1.92 smaller than Figures 1(a) and 1(b). Placement of Figure 1(c) on Figure 1(a), as has been done in Figure 1(d), demonstrates that the distant house on the left side is indeed equally large in both pictures. The house is also seen at the same distance in both pictures. Reducing house size increases perceived distance. The traffic signs are smaller and perceived at a longer distance in the small pictures. Equally perceived distances in Figures 1(c) and 1(d) show that picture size per se is probably irrelevant. Depicted object size seems the factor that determines perceived distance, not picture size. Changing size changes perceived distances but not depth. Figure 1(d) shows that depth between traffic signs and house is still compressed in the FL 72 mm picture relative to that in the FL 36 mm picture. Computations in the next paragraphs will reveal the distance information that characterizes depth.

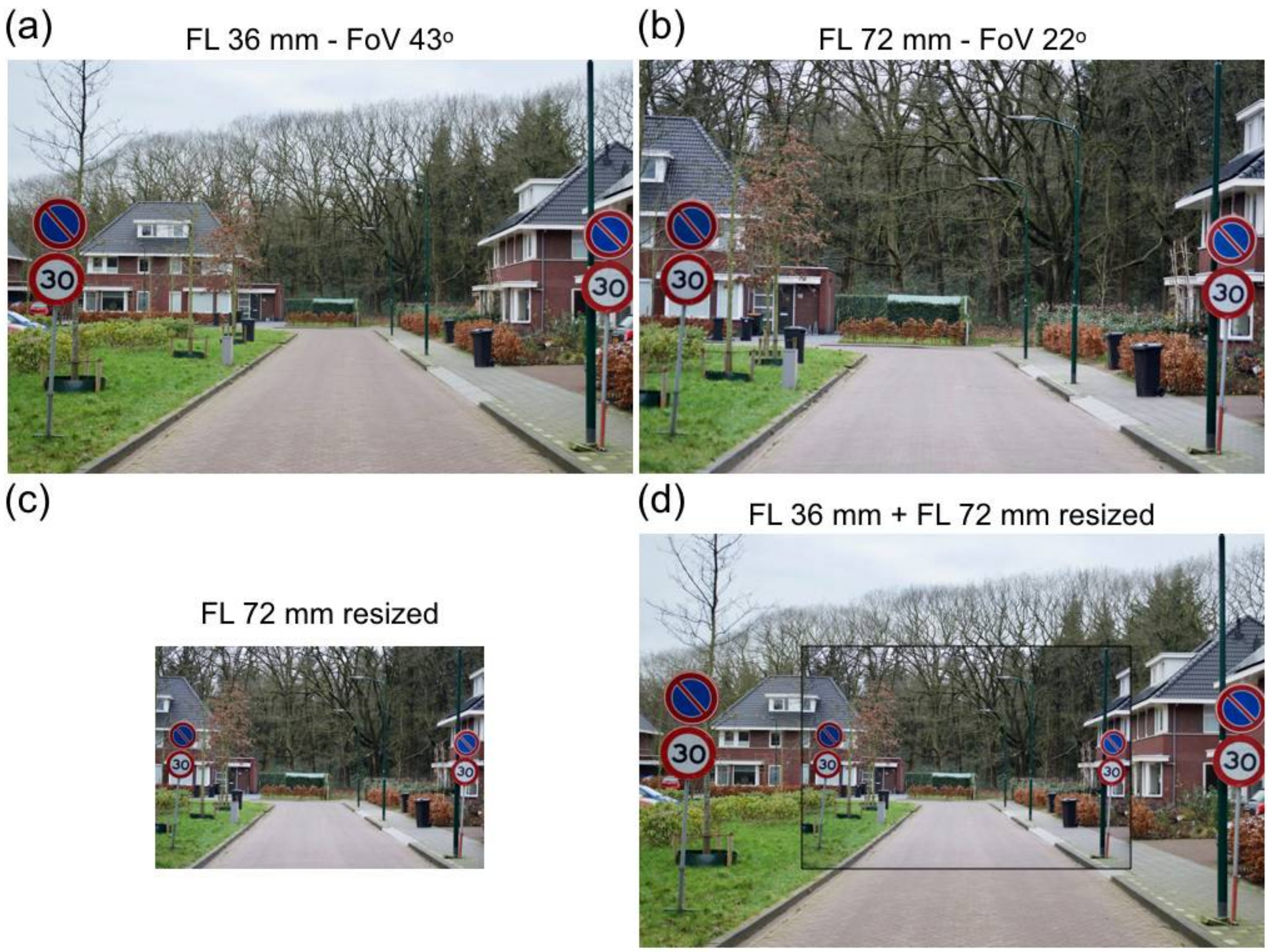

Figure 1. Two pictures of the same perspective scene. **(a)** The picture was taken with a camera equipped with a lens having a focal length (FL) of 36 mm and a field of view (FoV) of 43°. **(b)** The picture was taken with another lens (FL 72 mm, FoV 22°). **(c)** The picture is a resized version of Figure (b). The house on the left side is scaled to that of Figure (a). **(d)** Figure (c) including a thin black rim is placed on top of Figure (a).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at *Vision* **2018**, *2*, 26; doi:10.3390/vision2030026

3. Geometry of the physical scene

Distances of traffic signs and house relative to the camera positions were computed from the angular sizes of traffic signs and house in the pictures. Computations were made on pictures of Figures 1(a) and 1(b) measuring 8.9 x 13.5 cm. Knowledge of the FoVs of the two pictures and physical size of one familiar object was prerequisite for the unambiguous computation of the various distances. The 30 km traffic sign at the left side of the road, having a physical diameter of 60 cm, served as the familiar object in the computations. Figure 2(a) shows the distances computed from the pictures. The computed distances were verified by measuring the actual distances in physical space. Errors are smaller than 2%. The computed height of the house is 10.44 m while it actually is 10.23 m (data supplied by the builder of the houses). Figure 2(b) shows distances of traffic sign and house as function of camera position. The distances enable the computation of two measures of depth. Absolute depth is defined as the difference between distances of house and traffic sign. Relative depth is defined as the ratio between distances of house and traffic sign. Since traffic signs and house are stationary objects, absolute depth is independent of camera position (Figure 2(c)). Relative depth increases exponentially with more forward camera positions until it reaches infinity when the camera passes the traffic signs. The ratio is 4.4 at the position of the FL 72 mm picture and 8.1 at the position of the FL 36 mm picture. Distances and depths in physical space were computed in order to compare them to similar measures in pictorial space in the next paragraph.

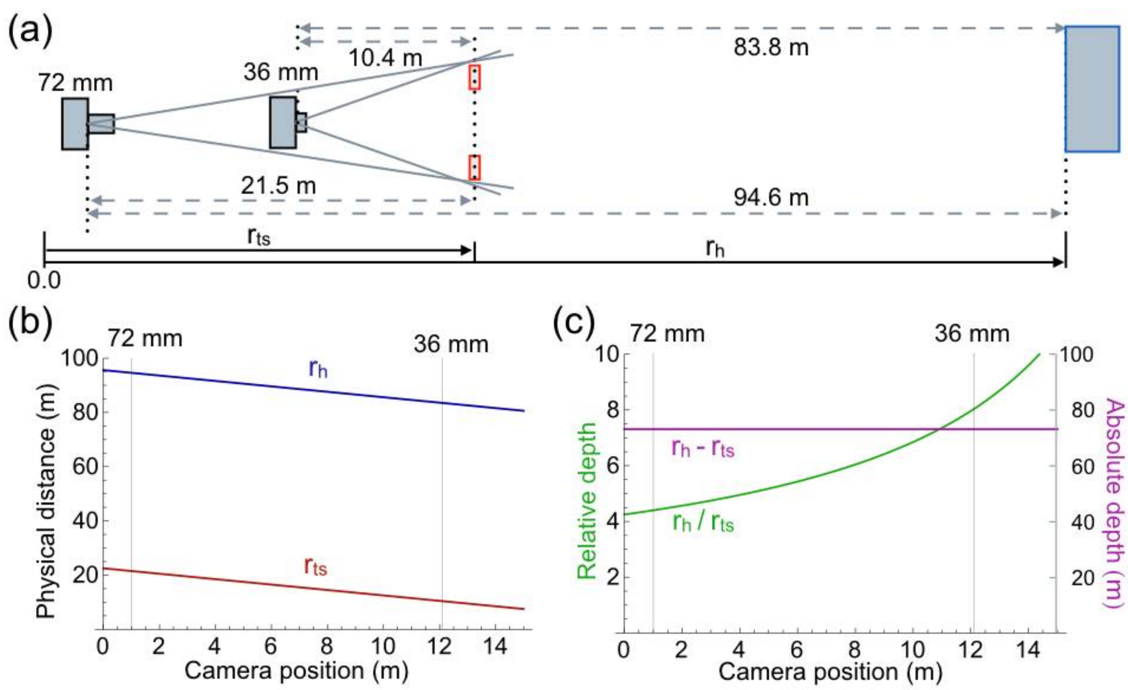

Figure 2. Geometry of the physical scene depicted in Figures 1(a) and 1(b). **(a)** Top view of the scene. Sizes and distances of objects are not to scale. Distances were computed of traffic signs (r_{ts}) and house (r_h) relative to the positions from which the pictures were taken. **(b)** Distance of traffic signs (r_{ts}) and house (r_h) as function of camera position. The arbitrary origin of the x-axis is chosen 1 m from the FL 72 mm camera position. Thin vertical lines mark the two camera positions from which the pictures were taken. **(c)** Relative (green) and absolute (magenta) depth between traffic signs and house computed as function of camera position.

4. Geometry of the pictorial scene

4.1. Distance

First a few observations are made about the perceived distances of objects in the pictures of Figures 1(a) and 1(b). The traffic signs are perceived at the same distance (r_{ts36} and r_{ts72}) in both pictures, irrespective of the viewing distance (r_p). Common properties of the traffic signs are equal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at *Vision* **2018**, *2*, 26; doi:10.3390/vision2030026

angular size in the pictures and equal size in physical space. The house is seen at the longest distance (r_{h36}) in Figure 1(a) and at a shorter distance (r_{h72}) in Figure 1(b). Depth between traffic sign and house is compressed in Figure 1(b) relative to Figure 1(a). The compression remains present for other viewing distances. For each viewing distance r_p, pictorial distances of traffic sign and house were computed from their size in physical space and their angular size in the pictures. Figure 3(a) shows qualitatively the geometry of the pictorial scene. Figure 3(b) shows that pictorial distances increase linearly with viewing distance r_p. The thin solid lines in Figure 3(b) indicate that pictorial distances are equal to physical distances at a viewing distance of 21 cm for the FL 36 mm picture, whereas this equality occurs at a viewing distance of 41 cm for the FL 72 mm picture. Depth is compressed for shorter viewing distances and expanded for longer viewing distances. The observation that the traffic signs are perceived at the same distance in both pictures shows that knowledge of camera positions and focal lengths are irrelevant for pictorial distance. Pictorial distance is determined by the angular size of the depicted object and knowledge of the object's size in physical space. Figure 3(c) shows computations of depth. Absolute depth increases linearly with viewing distance (Figure 3(c)). The slope is 48% steeper for the FL 36 mm picture. Relative depth is constant as function of viewing distance. The ratio is 8.1 for the FL 36 mm picture and 4.4 for the FL 72 mm picture, which is a difference of 42%. Both types of depth seem consistent with the compression of depth perceived in the FL 72 mm picture of Figure 1(b) relative to that in the FL 36 mm picture of Figure 1(a), independent of viewing distance.

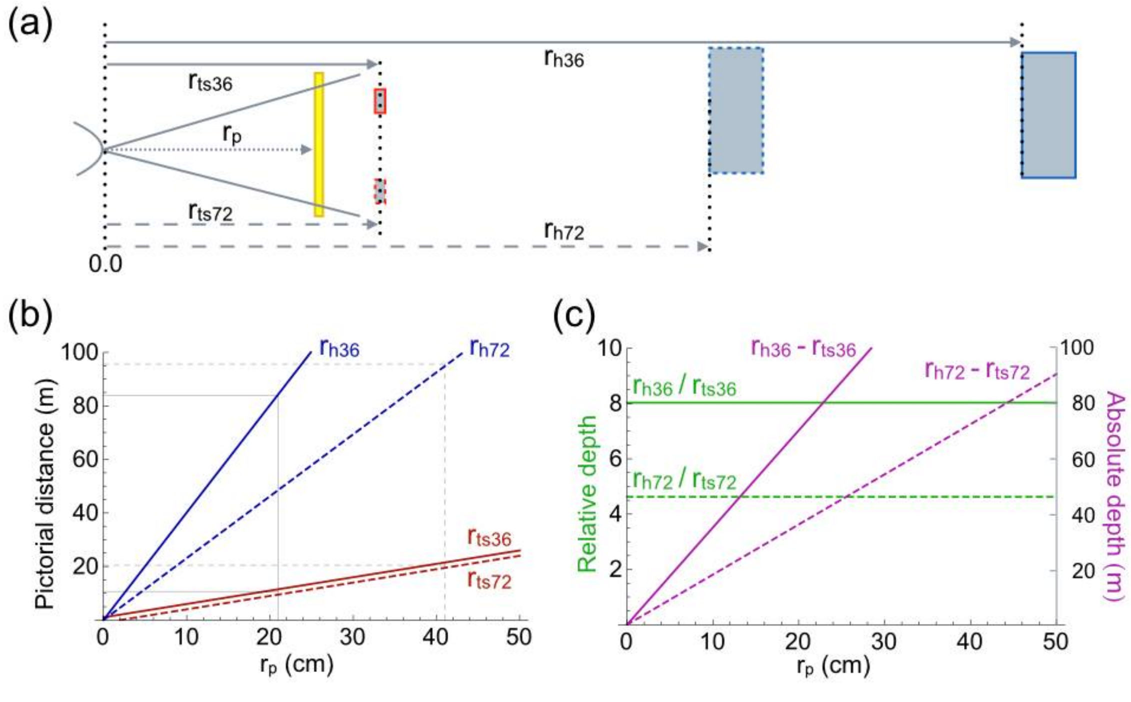

Figure 3. Geometry of the pictorial scenes of Figures 1(a) and 1(b). **(a)** Top view of the pictorial scene. Sizes and distances of objects are not to scale. Distances were computed for the traffic signs (r_{ts36} and r_{ts72}) and house (r_{h36} and r_{h72}) depicted in the two pictures. The viewer (curve at the left side) is looking at the pictures (yellow bar) from a distance r_p. **(b)** Pictorial distances of traffic signs and houses that produce the same retinal images as their projections in the pictures. Distances are computed as function of r_p. Thin lines mark the viewing distances at which the computed distances are equal to the physical distances (see Figure 2(a)). **(c)** Relative (green) and absolute (magenta) depth between the computed distances of traffic signs and houses as function of r_p.

Further evidence for the thesis that angular size specifies pictorial distance comes from the small pictures of Figures 1(c) and Figure 1(d) where Figure 1(c) has been placed on top of Figure 1(a). Both traffic sign and house appear further away in Figures 1(c) and 1(d) than in Figure 1(b). The house in Figures 1(a) and 1(d) appears at the same distance. Figure 4(a) shows qualitatively the geometry of the pictorial scene of Figure 1(d). Figures 4(b) and 4(c) show the computed pictorial distances. Figure

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at Vision 2018, 2, 26; doi:10.3390/vision2030026

4(b) shows that the two traffic signs have very different distances now, whereas the distances of the houses are indeed equal. The equal distances of the house show again that pictorial distance is determined by the angular size of the depicted object and knowledge of the object's size in physical space. The thin solid lines in Figure 4(b) indicate that pictorial distances are equal to physical distances at a viewing distance of 21 cm for the FL 36 mm picture, whereas this equality occurs at a viewing distance of 24 cm for the small FL 72 mm picture. The difference between the two viewing distances reflects the fact that the two pictures were taken from different camera positions (Figure 2(a)). Resizing Figure 1(b) to Figure 1(c) changes absolute depth but not relative depth (compare Figures 4(c) and 3(c)). Absolute depth increases linearly with viewing distance again (Figure 4(c)). Reduction in size of the FL 72 mm picture has increased absolute depth to a level that is approaches absolute depth in the FL 36 mm picture. The difference in slope is reduced to just 12%. Reduction in size of the FL 72 mm picture does not have any effect on relative depth. Relative depth is independent of picture size. The independence of picture size seems compatible with the depth compression that is perceived between house and traffic sign in the large (Figure 1(b)) and small (Figure 1(d)) FL 72 mm pictures. The next paragraph explores how relative depth is related to angular size.

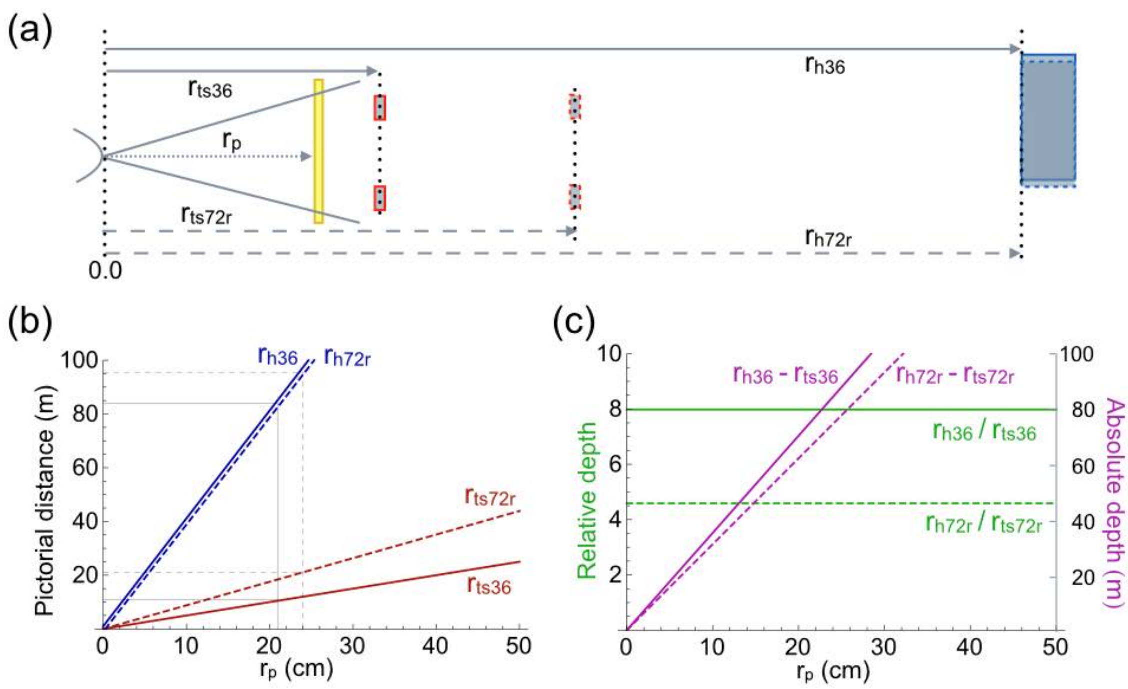

Figure 4. Geometry of the pictorial scenes of Figure 1(d). **(a)** Top view of the pictorial scene. Sizes and distances of objects are not to scale. Distances were computed for the traffic sign (r_{ts36} and r_{ts72r}) and house (r_{h36} and r_{h72r}) depicted in the two pictures. The viewer (curve at the left side) is looking at the pictures (yellow bar) from a distance r_p. **(b)** Pictorial distances of traffic signs and houses that produce the same retinal images as their projections in the pictures. Distances are computed as function of r_p. Thin lines mark the viewing distances at which the computed distances are equal to the physical distances (see Figure 2(a)). **(c)** Relative (green) and absolute (magenta) depth between the computed distances of traffic signs and houses as function of r_p. .

4.2. Angular size

Computation of angular size is relevant because of the demonstrated relationship to perceived distance in pictures. Figure 5(a) shows angular sizes of the physical traffic sign and house as function of camera position. Angular size of the traffic sign shows much more variation than that of the house due to the much shorter distance between traffic sign and camera. Variations in the angular sizes are much more similar in pictorial space (Figure 5(b)) due to almost equal distances of the depicted traffic sign and house to the viewer. Angular size ratio, defined as angular size of the house divided by angular size of the traffic sign, was computed to demonstrate the differences between variations in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at Vision 2018, 2, 26; doi:10.3390/vision2030026

detail. In physical space, angular size ratio decreases with more forward positions of the camera because the traffic sign increases much faster in size than the house does (Figure 5(c)). In pictorial space, angular size ratio remains constant for viewing distances longer than 10 cm because the depicted traffic sign and house are equally distant from the viewer (Figure 5(d)). Magnitude of the constant is determined by the angular sizes of traffic sign and house at the positions from which the photographs are taken (Figure 5(a)). Focal length of the lens as such does not affect the angular size ratios because changes of focal length amplify the angular sizes of traffic sign and house by identical factors.

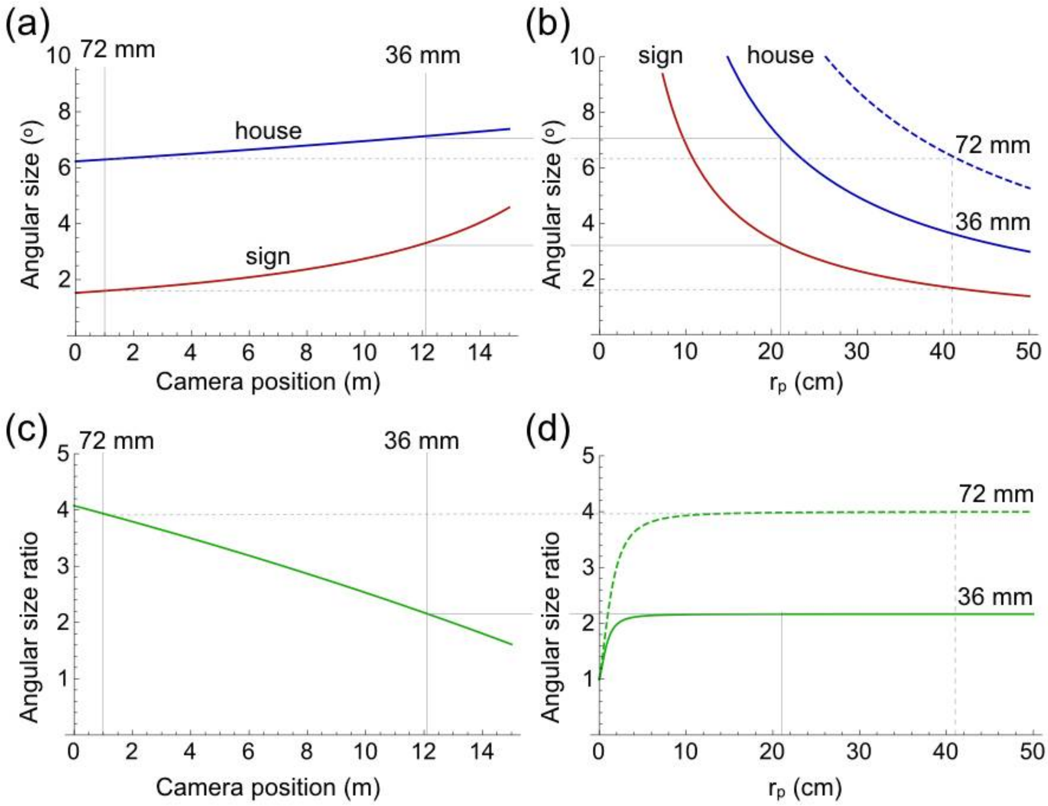

Figure 5. Angular size. **(a)** Angular sizes of the physical traffic sign (red) and house (blue) as function of camera position. The origin of the camera-position scale is given in Figure 2(a). **(b)** Angular sizes of the depicted traffic sign (red) and house (blue) as function of viewing distance of the picture (r_p). The solid blue line shows the angular size of the house in the 36 mm picture of Figure 1(a), the dashed blue line that of the house in the 72 mm picture of Figure 1(b). **(c)** Ratio between angular sizes of the physical house and traffic signs shown in (a). **(d)** Ratio between angular sizes of the depicted house and traffic signs shown in (b). The horizontal solid line shows the ratio in the 36 mm picture and the dashed line in the 72 mm picture.

4.3. Visual direction

For a good comparison of pictorial space and physical space, it is of importance to compare visual directions in the two spaces. Figure 6(a) shows visual directions for two physical objects, one nearby and the other far away, as function of camera position. Computations were made for the centre of the 30 km traffic sign standing at the left side of the road and the mailbox of the distant house in Figures 1(a) and 1(b). Visual directions of traffic sign and mailbox change by very different amounts between the long-lens (FL 72 mm) and short-lens (FL 36 mm) camera positions. Due to the short object to camera distance, visual direction of the traffic sign doubles in eccentricity. Visual direction of the mailbox hardly changes because of the long distance between house and camera. Figure 6(b) shows the visual directions for the centre of the depicted traffic sign and mailbox as function of viewing distance. Due to identical distances between the depicted objects and viewer, visual directions of traffic sign and mailbox change by equal factors between two the viewing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at Vision 2018, 2, 26; doi:10.3390/vision2030026

distances at which the pictures of Figures 1(a) and 1(b) were taken. Between the 21 cm and 41 cm viewing distances all visual directions become less eccentric by a factor of two. Visual direction ratio, defined as visual direction of the house divided by angular size of the traffic sign, was computed for the physical and depicted objects. In physical space, visual direction ratio decreases with more forward positions of the camera because the traffic sign becomes much faster eccentric than the house does (Figure 5(c)). In pictorial space, visual direction ratio remains constant for viewing distances longer than about 10 cm because the depicted traffic sign and house are equally distant from the viewer (Figure 5(d)). Visual directions of the physical traffic sign and house at the positions from which the photographs are taken (Figure 5(a)) determine the values of the ratio. Like for angular size, focal length of the lens does not affect the visual direction ratios because changes of focal length amplify the visual directions of traffic sign and house by identical factors.

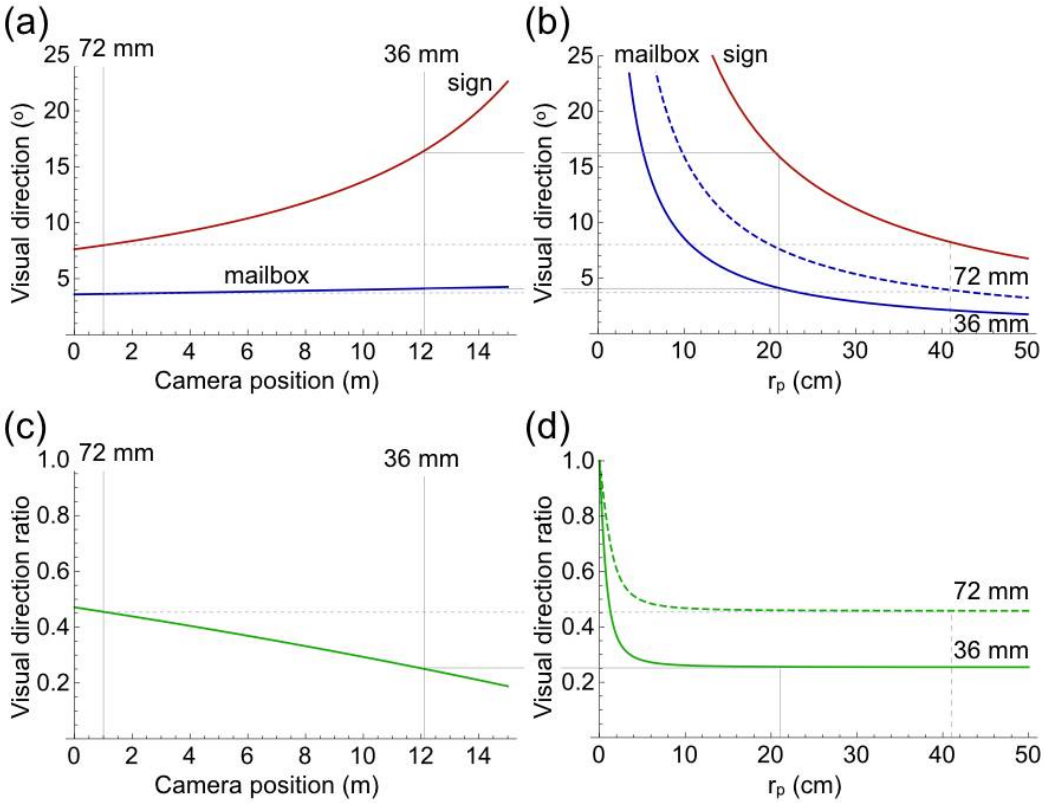

Figure 6. Visual directions. **(a)** Visual directions of the centre of the physical 30 km traffic sign standing at the left side of the road (red) and the mailbox of the distant house (blue) as function of camera position. The origin of the camera-position scale is given in Figure 2(a). **(b)** Visual directions as function of viewing distance of the picture (r_P). The solid blue line shows the visual direction of the mailbox in the 36 mm picture of Figure 1(a), the dashed blue line that of the mailbox in the 72 mm picture of Figure 1(b). Thin lines mark the viewing distances at which the visual directions of projected objects are equal to the visual directions of the physical objects (see Figure 2(a)). **(c)** Ratio between visual directions of the physical house and traffic signs shown in (a). **(d)** Ratio between visual directions of the depicted house and traffic signs shown in (b). The solid line shows the ratio in the 36 mm picture and the dashed line in the 72 mm picture.

5. Sizes, distances and distortions in pictures

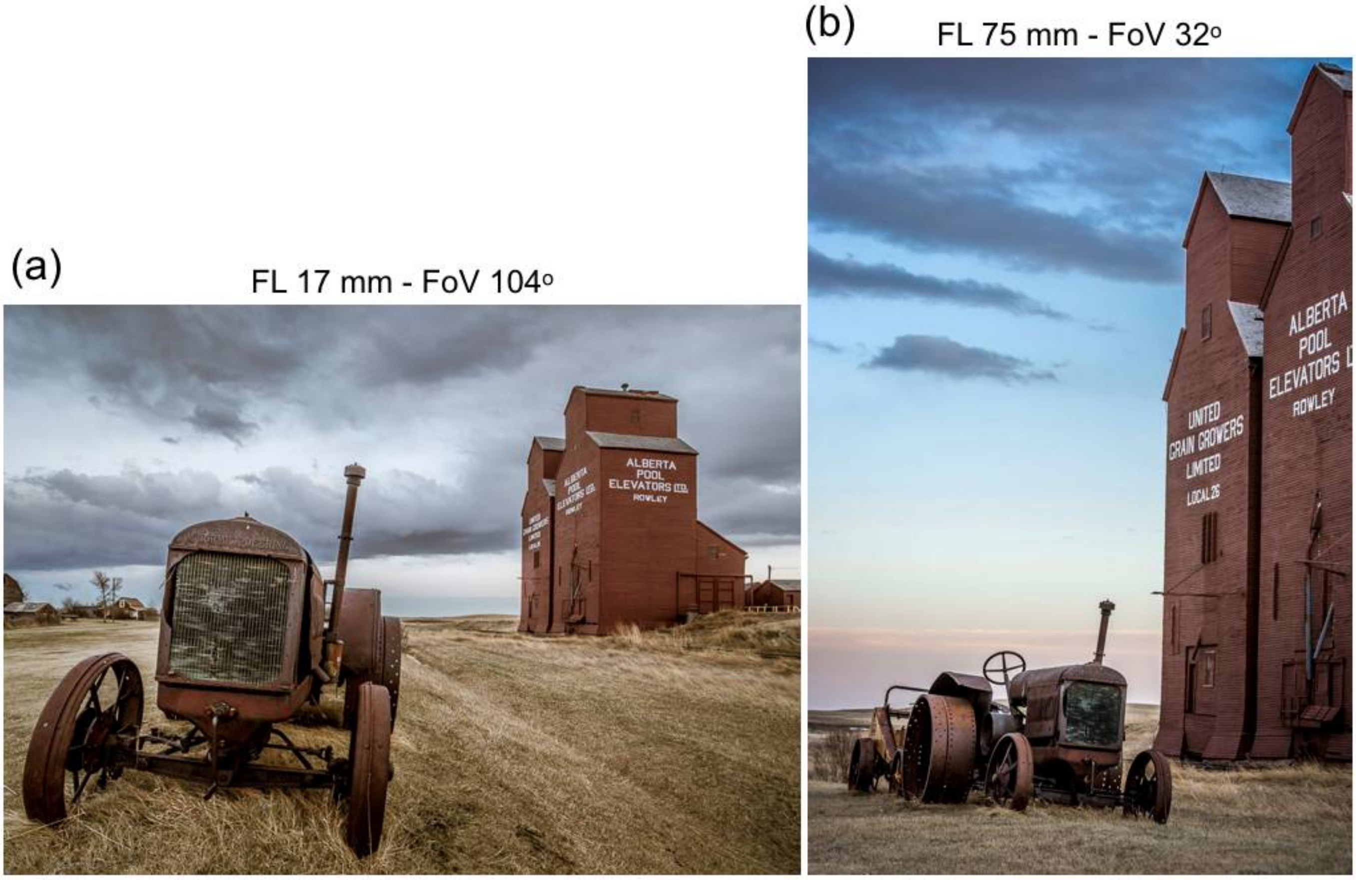

Figure 7. Two pictures of the same perspective scene. Darlene Hildebrandt, a professional photographer, took both pictures (https://www.digitalphotomentor.com/5-mistakes-beginners-make-using-a-wide-angle-lens-and-how-to-avoid-them/). **(a)** The picture was taken with a full-frame camera equipped with a short lens having a focal length of 17 mm and a field of view of 104°. **(b)** This picture was taken with a long lens (FL 75 mm, FoV 32°). Pictures were used for this study after written consent of the photographer.

Figure 7 shows two pictures taken by a professional photographer of a tractor and grain elevator. At first sight, sizes and distances of both objects seem incompatible in the two pictures. The picture of Figure 7(a) was taken with a short lens having a very wide FoV, whereas the picture of Figure 7(b) was taken with a long lens having a narrow FoV. On the basis of the familiar object sizes and the objects' positions in the picture, the tractor is seen nearby and the grain elevator rather far away in the left picture. The grain elevator appears much nearer in the right picture. It is as if the photographer has moved forward. The distance between tractor and elevator seems much shorter in the right picture. If so, then the tractor must have been moved closer to the grain elevator. According to the photographer (see her website), however, the tractor was at the same position in both pictures (on the basis of the tractor's looks, moving it may not be easy). The immobility of the scene, the FoVs and sizes of the pictures, and the physical size of one object make it possible to reconstruct the geometry of the physical scene and the positions from which the pictures were taken (Figure 8(a). The track width of 165 cm (65 inch) of the classic Mc Cormick Deering 22-36 tractor served as the familiar size of one physical object. Figure 8(a) shows that the left picture of Figure 7 is taken from a far longer distance than the right picture. As a naïve viewer, one has no idea of the different camera positions and settings. As the previous analyses of this study showed perceived distances are signalled by the angular size of pictured objects. Even knowledge of position and settings of the camera, as the professional photographer may have, does not help in seeing objects at physically correct sizes and distances. Figure 8(b) shows that resizing the pictures, so that the grain elevators

doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at *Vision* **2018**, *2*, 26; doi:10.3390/vision2030026

are depicted on the same scale, makes that distances of the pictured objects are better in line with physical distances. The tractor on the small picture is seen at a much farther distance than the tractor on the large picture. The grain elevator in the small picture is seen slightly farther away than the grain elevator in the large picture due to the different camera orientations. The pictorial distances are now qualitatively in accordance with the different camera positions and geometry of the physical scene (Figure 8(a)).

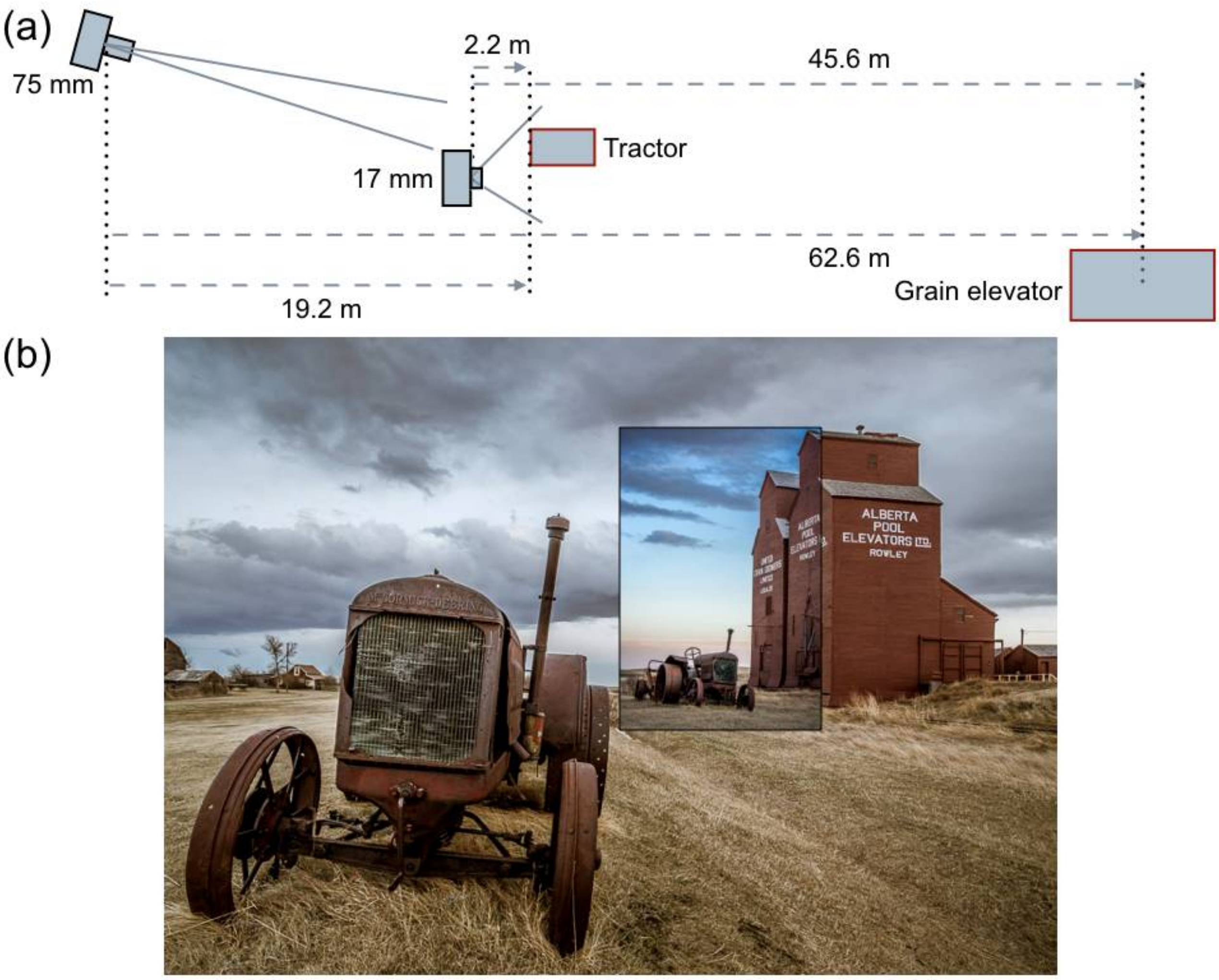

Figure 8. (a) Geometry of the physical scene. **(b)** Figure 7(a) is shown here with Figure 7(b) on top. Figure 7(b) has been resized such that the grain elevator is depicted at the same scale in both pictures.

The two tractors shown in Figure 8(b) seem to have different shapes. A short-focal-length lens has been used for the nearby tractor and a long-focal-length lens for the far one. Objects captured with short lenses appear expanded in depth, while those captured with long lenses appear compressed. Figure 1 and the following analysis shows that depth expansion and compression are not related to the length of lenses but angular size of the depicted objects. Depth compression and expansion can also affect the appearance of a face. Long lenses make a person look smarter, more attractive, and less approachable; short lenses have the opposite effects (Perona, 2007). The effect of lenses on the appearance of a face was examined by comparing photographs taken with long and short lenses from different distances. Figure 9(a) shows the face of a bronze statue of a girl photographed with a short lens of 18 mm. Distance between camera and forehead was 0.15 m. Figure 9(b) shows the face with a longer lens of 55 mm. The distance between camera and forehead was 0.80 m now. The 55 mm lens mounted on an APS-C camera is within the range of focal lengths that is often used for photographing faces. Indeed, the girl's face looks more natural and attractive. Figure 9(c) shows again the girl photographed with the short lens of 18 mm, but now with a camera distance

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at Vision 2018, 2, 26; doi:10.3390/vision2030026

of 0.80 cm from the forehead. Due to the longer distance, a larger part of the girl than her head is visible on the picture. Figure 9(d) shows a cropped version of Figure 9(c). Comparison of Figures 9(b) and 9(d) shows that the girl's head looks very similar in both pictures, irrespective of the different lenses used for both pictures. Contrastingly, the girl's head looks different in Figures 9(a) and 9(d) although the same short lens was used for both pictures. The pictures of Figure 9 show that lenses do not cause facial distortions. It is the farther camera position that determines the attractiveness of faces. Also the physical face of the bronze girl looks as is shown in Figure 9(a) if the viewer is just 0.15 m away from the girl's head. For instance, one cannot see the topside of the head from that position. Due to the three-dimensional shape of the head, a larger portion becomes visible and relative distances between different parts of the face become smaller if one views the statue from a more distant position.

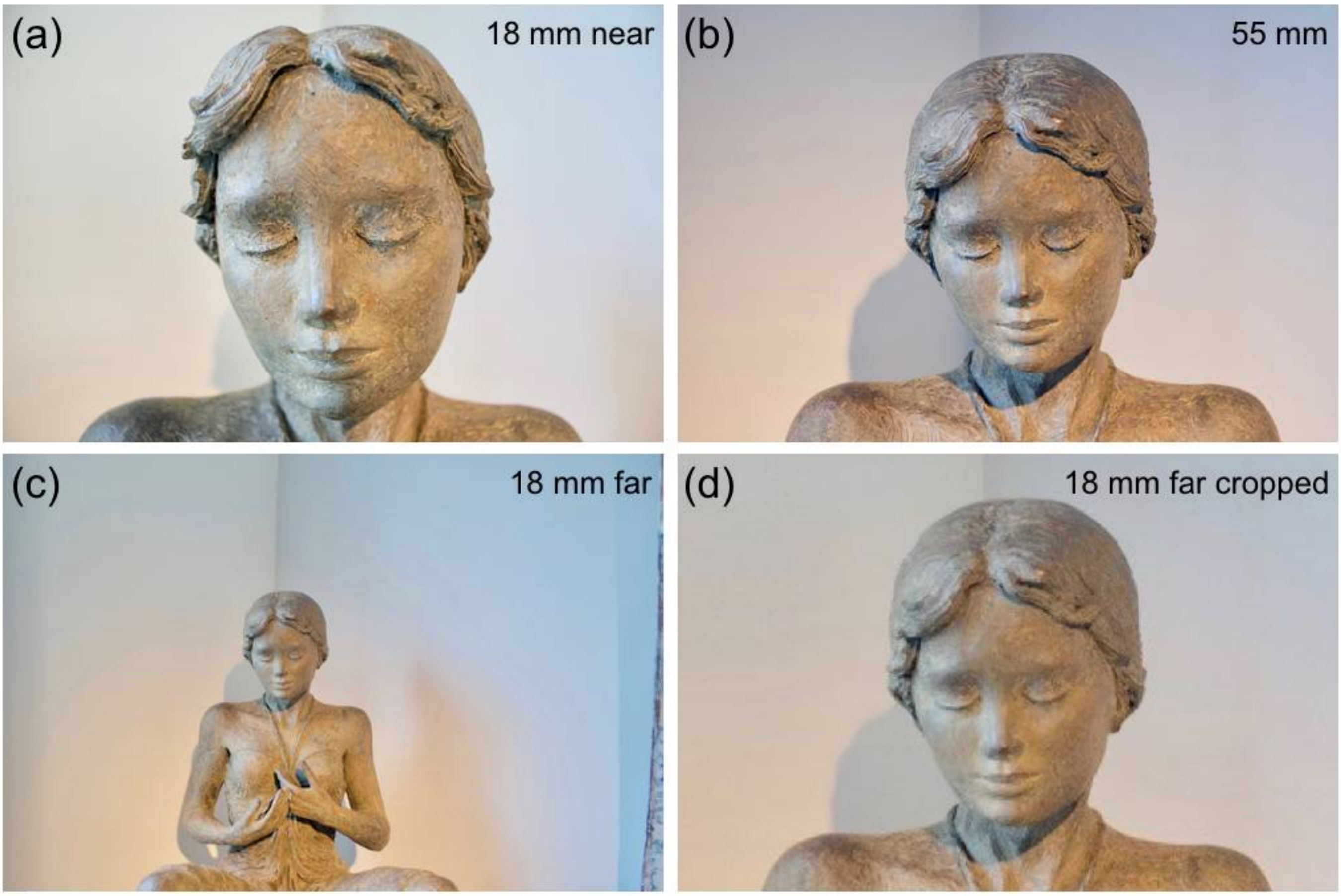

Figure 9. Four pictures of a bronze sculpture of a girl. The camera is fitted with an APS-C sensor. **(a)** Picture taken with an 18 mm lens from a near camera position. **(b)** Picture taken with a 55 mm lens from a far camera position. **(c)** Picture taken with an 18 mm lens from the far camera position of (b). **(d)** Picture (c) but cropped and expanded to match the size of the face in the other pictures.

6. Distance in pictorial space fitted with perspective distance functions

Until now the geometry of pictorial space was compared to that of physical space and not visual space. Yet, it has been documented exhaustively in the literature that visual space differs from physical space, in the depth domain. Recent analysis showed that perspective space is a good model of visual space (Erkelens, 2015c). It is simple yet powerful model because it describes many experimental results, explains certain visual phenomena and unifies a number of models of distance perception (Erkelens, 2017). To investigate resemblance between distance in pictorial space and visual space, the literature was searched for studies that provided experimental results of depth perceived in pictures. Kraft and Green (1989) presented an extensive set of data, which has already been re-analyzed by Cutting (2003). The present finding that angular size is key to distance perception

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at *Vision* **2018**, *2*, 26; doi:10.3390/vision2030026

in pictures warrants a third analysis of Kraft and Green's data. Kraft and Green (1989) collected data of 70 observers who judged the distance of objects in pictures. Kraft and Green presented many photographs. Photographs were taken with a full-frame camera and five different lenses: focal lengths of 17, 35, 48, 75, and 135 mm. With such lenses the horizontal field of view subtends about 105, 60, 45, 32, and 20°, respectively. In two different outdoor environments, Kraft and Green (1989) planted poles at distances of 20, 40, 80, 160, and 320 m from a fixed camera. Viewing different arrangements of 50 slides, observers made judgments of the distance of each pole from the camera. Figure 10(a) shows the graph copied from Kraft and Green (1989). The graph shows that perceived distance depends strongly on the length of the lens. In view of the result of this study that angular size is the effective stimulus for distance perception it is relevant to note that Kraft and Green (1989) projected all photographs with a Kodak Carousel slide projector on a screen. All pictures had one size implying that angular sizes of poles in long-lens pictures were too wide relative to those of poles in short-lens pictures. To correct for the different FoVs, physical distances of poles to the camera were renormalized to distances as if the photographs were taken with the FoV of a FL 50 mm lens. Figure 10(b) shows that the majority of data form a single curve after renormalization. Perceived distances of poles in the narrow field pictures of 17 mm deviate from the main curve for the longer physical distances of the poles. The data were fitted with the following distance function for perspective space (Erkelens, 2017):

Perceived distance = vd x Physical distance / (vd + Physical distance), where *vd* indicates distance of the vanishing point of perspective space. Figure 10(b) shows two separate fits, one to the 17 mm data and the other fit to the data of the other FLs. Both fits account for more that 99% of the variance. A single fit to all data accounts for just slightly less of the variance, namely 98% (Figure 10(c)). The excellent fits show that perspective space is a good model for pictorial distance. There is no clear argument that explains the difference between the 17 mm and other FLs in the study of Kraft and Green (1989).

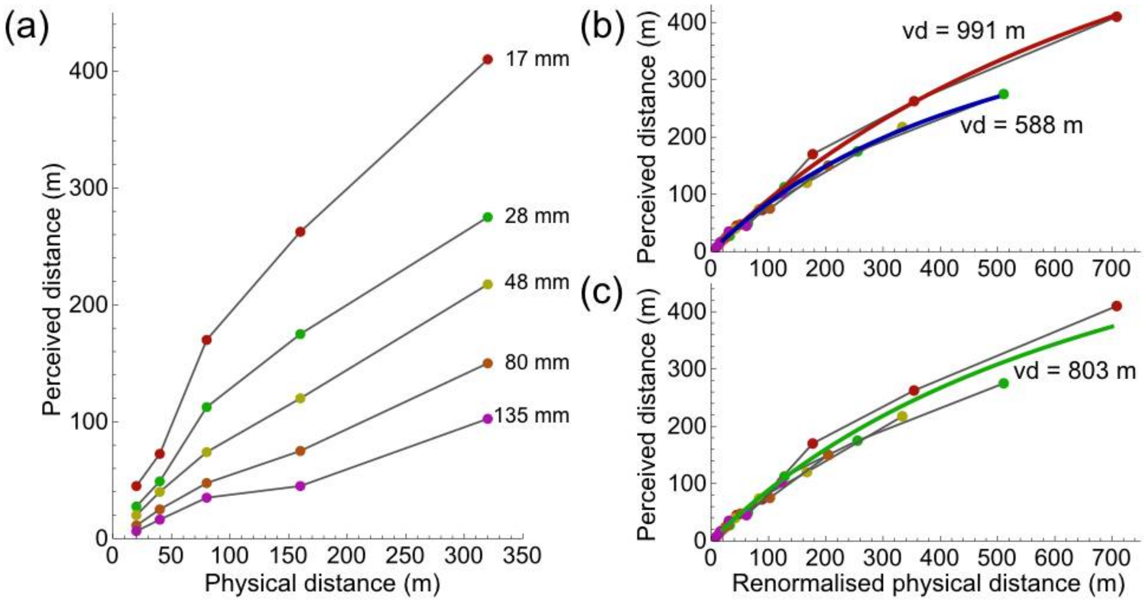

Figure 10. Perceived distance in pictures. **(a)** Data copied from Kraft and Green (1989). **(b)** Data of (a) with physical distances renormalized to the FoV of a FL 50 mm lens. Separate fits were computed for perspective distance curves for the 17 mm data (red) and the other data (blue). **(c)** Data of (b) fitted with a single perspective distance curve (green).

7. Discussion

The straightforward computations of this study provide insight in the geometry of depicted scenes and the pictorial space that we experience when looking at pictures. Picture size and FoV of the lens are necessary data for the veridical reconstruction of directions and distances of physical objects relative to the camera position. Figure 1 shows traffic signs whose physical distances (Figure 2) were considerably different in Figures 1(a) and 1(b). The fact that we perceive the equally large traffic signs at equal depths in both pictures shows that pictorial distances are judged relative to the viewer rather than camera position. The current study shows that angular size determines the pictorial distance of familiar objects. This conclusion explains previous research showing that viewers do not compensate for incorrect viewing distance (Banks, Cooper & Piazza, 2014; Bengston, Stergios, Ward, & Jester, 1980; Cooper, Piazza & Banks, 2012; Kraft & Greene, 1989; Smith & Gruber, 1958; Todorovic, 2009). FoV of the lens is irrelevant for the computation of pictorial distances. Picture size does neither appear in the computations. The irrelevance of picture size is easily verified by covering a part of a picture in Figures 1 or 5. Picture size only affects perceived distance if it changes the angular size of depicted objects. This occurs when pictures are uniformly compressed or enlarged. Angular size is the effective stimulus for perceived size and distance in the perspective space model (Erkelens, 2017). The model proved to be a successful model for the visual perception of sizes, distances and angles of real objects in physical space (Gilinsky, 1951, 1955; Erkelens, 2015, 2017). Reanalysis of the data of Kraft and Green (1989), such that the perceived distances of depicted objects were related to angular size, shows that perspective space is also a good model for pictorial distances. The distances of vanishing point of the model fits shown in Figures 10(b) and 10(c) may not represent the real vanishing distance of pictorial space because it is not clear from Kraft and Green's paper whether normalization of the data to the FL 50 mm format produced the correct angular sizes or angular sizes times a magnification factor.

7.1. Invariance in picture perception

The visual scene changes considerably if an observer moves forward in a physical environment with nearby and far objects. Adjacent objects approach faster and move outward in front of far objects until they disappear out of sight. Relative distances, directions and sizes remain only constant if the observer stands still. Scene dynamics are very different if you move toward a picture of the same environment. Although all objects move outward and get closer, the scene itself is frozen. Absolute distances, directions and sizes of the depicted objects depend on viewing distance. However, as this study shows, relative distances, directions and sizes remain constant and are associated with standstill in a real three-dimensional environment. The relationship with immobility may explain the invariance of pictorial scenes if these are viewed at incorrect distances. Scenes are considered trustworthy representations of physical scenes over a long range of viewing distances before they seem compressed or expanded in extreme conditions. Estate agents exploit the invariance by presenting rooms wider than they really are in their advertisements. Perceptual invariance across viewing position has been of interest to many visual scientists. Most studies investigated the shape of pictorial objects from oblique viewing positions (Cutting, 1987; Erkelens, 2013a, 2013b; Hagen, 1976; Goldstein, 1987; Koenderink et al., 2004; Rogers, 1995; Rosinski & Farber, 1980; Vishwanath et al., 2005). Effects of viewing distance on perceived size, distance and direction has received less attention (Cooper et al., 2002; Banks et al., 2014). Apparently, perceptual invariance as function of viewing distance has usually been taken for granted.

7.2. Misconceptions about picture perception

There exists a widely held erroneous belief in the literature of picture perception. The belief is that pictures viewed from positions other than the camera position are valid two-dimensional representations of physical space. The consequence of the belief is that perceived distances and directions obtained from such pictures can be used to draw conclusions about the geometry of pictorial space. Although the belief is usually not expressed explicitly, there are various examples of incorrect conclusions. One incorrect conclusion involves pictorial distortions. At a certain viewing distance, called the correct viewing distance, a picture induces the same retinal image as the physical scene. A picture is viewed from the correct distance if the angular size of the picture equals the FoV of the lens. Changing the size of a picture by compression or expansion changes the correct viewing distance too. Figures 1, 3 and 4 illustrate this. Compression of Figure 1(b) to 1(c) changed the correct

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at Vision 2018, 2, 26; doi:10.3390/vision2030026

viewing distance from 41 cm (Figure 3(b)) to 22 cm (Figure 4(b)). Viewing the picture from an incorrect distance magnifies the angular sizes of all depicted objects by a common factor (Figure 5(d)). Compression or expansion of the picture causes the same effect. Such changes in angular size are very different from those associated with moving forward or backward in the physical environment. Then, near objects change much more in angular size than far objects do (Figure 5(a)). Due to the angular size – distance relationship, pictorial distances behave different as function of viewing distance than visual distances as function of camera position (compare Figures 3 and 4 with Figure 2). A widely held opinion is that photographs of scenes captured with short-focal-length lenses appear expanded in depth, while those captured with long lenses appear compressed (Cooper et al., 2012). The examples presented in Figures 1, 8 and 9 show that expansion or compression of depth is not related to focal length but to the angular sizes of nearby and far objects in the picture. The ratio between angular sizes of depicted objects depends on the camera position from which the picture is taken, not on focal length of the lens. Each ratio is unique for a certain camera position (Figure 5(c)), implying that the picture could not have been made from any other camera position. From an incorrect viewing distance, the picture is a projection of a non-existing physical scene. Therefore, it is inappropriate to call depth expansion or compression distortions because, from incorrect viewing distances, they reflect the correct perspective projection of non-existing physical scenes. A related misconception concerns the distortion of faces by short-focal-length lenses (Banks et al., 2014; Cooper et al., 2012). Short-focal-length lenses do not expand depth if pictures are viewed at the correct distance. The example of the bronze girl in Figures 9(c) and 9(d) shows that such lenses not necessarily exaggerate the depth of a face. Facial distortions occur when pictures that were taken from extremely close camera positions are viewed from longer distances. Another incorrect conclusion involves pictorial direction. Koenderink, van Doorn, de Ridder and Oomes (2010) have argued that visual directions are parallel. Evidence came among others from perceptual judgments of the people's orientations while these were seated on chairs next to each other with ample space between them. Naïve observers made the judgments from pictures taken by a camera equipped with wide-angle lenses (horizontal FoV 104°). A linear-perspective picture showed the persons, i.e. the authors, in a fronto-parallel row. The authors' orientations were judged as rotated with respect to the straight-ahead direction. An equiangular projection showed the authors in a circular (about the camera) row, all facing the camera. Now the authors were judged as fronto-parallel and seated in strict military order. The conclusion that visual directions are parallel denies the fact that the visual and pictorial spaces are perspective of nature. The tractor of Figure 7(a) provides a good illustration. The left front wheel of the tractor is oriented to the camera and positioned close to the centre of the picture. Figure 7(b) shows that the front wheels of the tractor are parallel to each other in physical space. Therefore, both wheels are aligned with the straight-ahead direction of the camera in Figure 7(a). At the off-centre position of the right front wheel, the wheel is not oriented to the camera in physical space. If we would look directly at the right front wheel from the same position we would see it at an angle too. To join Koenderink's military terminology, in a platoon off-centre soldiers do not appear to look at their commander. They better do not! This leaves unanswered the question why naïve observers judged the facing authors as fronto-parallel. The answer is probably related to the fact that judgments were made from pictures and the observation that wide-angle pictures are usually viewed from too far (Cooper et al., 2012). The tractor of Figure 7(a) is again a good illustration. From the geometrical data of Figure 8(a) and the known track width of the front wheels it was computed that the physical right front wheel makes an angle of 37° with its direction to the camera. Prolonging the viewing distance or uniformly compressing a picture leaves the shape of depicted objects unchanged. However, if the picture is viewed from 20 cm in front of the left front wheel, the visual angle to the right front wheel is only 17° instead of 37°. As a result the right front wheel will appear rotated outward 20°. The generalization from this example is that depicted objects will usually appear rotated outwards relative to the physical orientations in short-focal-lens pictures.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0289.v1

Peer-reviewed version available at *Vision* **2018**, *2*, 26; doi:10.3390/vision2030026

8. Conclusions

Computations made on different pictures of one perspective scene reveal that angular size is the effective stimulus for the perceived distance of objects in pictures. Although pictorial distances and directions of object change as function of viewing distance, ratios of distances and directions are constant. Pictorial distances and directions were computed from pictures by using the rules that predict visual distances and directions of physical objects. Data of distance judgments obtained from the literature shows that perspective space is as good a model for pictorial space as it is for visual space. The derived pictorial geometry reveals a few misconceptions about picture perception.

References

1. Baird, J. C., & Wagner, M. (1991). Transformation theory of size judgment. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 852-864.
2. Banks M. S., Cooper E. A., & Piazza E. A. (2014). Camera focal length and the perception of pictures. *Ecological Psychology, 26(1-2)*, 30–46.
3. Bengston, J. K., Stergios, J. C., Ward, J. L., & Jester, R. E. (1980). Optic array determinants of apparent distance and size in pictures. *Journal of Experimental Psychology: Human Perception & Performance, 6*, 751–759.
4. Bian, Z., Braunstein, M. L., & Andersen, G. J. (2005). The ground dominance effect in the perception of 3-D layout. *Perception & Psychophysics, 67*, 815–828.
5. Busey, T. A., Brady, N. P., & Cutting, J. E. (1990). Compensation is unnecessary for the perception of faces in slanted pictures. *Perception & Psychophysics, 48(1)*, 1–11.
6. Cooper E. A., Piazza E. A., & Banks M. S. (2012). The perceptual basis of common photographic practice. *Journal of Vision, 12(5),* 8, 1–14.
7. Costall, A. P. (1990). Seeing through pictures. *Word & Image, 6*, 273–277.
8. Cutting, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. *Journal of Experimental Psychology: Human Perception and Performance 13*, 323–334.
9. Cutting, J. E. (2003). Reconceiving perceptual space. In: *Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space* (ed. Hecht, H., Schwartz, R. & Atherton, M.), 215–238. Cambridge, MA: MIT Press.
10. Da Silva, J.A. (1985). Scales for perceived egocentric distance in a large open field: Comparison of three psychophysical methods. *American Journal of Psychology, 98,* 119–144.
11. Erkelens, C. J. (2013a). Virtual slant explains perceived slant, distortion and motion in pictorial scenes. *Perception, 42*, 253–270.
12. Erkelens, C. J. (2013b). Computation and measurement of slant specified by linear perspective. *Journal of Vision, 13*(13), 16, 1–11.
13. Erkelens, C. J. (2015a). The extent of visual space inferred from perspective angles. *i-Perception, 6(1)*, 5-14.
14. Erkelens, C. J. (2015b). Perception of perspective angles. *i-Perception, 6(3)*, 1–11.
15. Erkelens, C. J. (2015c). The perspective structure of visual space. *i-Perception, 6*(5), 1–13.
16. Erkelens, C. J. (2017). Perspective space as a model for distance and size perception. *i-Perception, 8*(6), 1–20.
17. Feria, C. S., Braunstein, M. L., & Andersen, G. J. (2003). Judging distance across texture discontinuities. *Perception, 32*, 1423–1440.
18. Foley, J. M., Ribeiro-Filho, N. P., & Da Silva, J. A. (2004). Visual perception of extent and the geometry of visual space. *Vision Research, 44*, 147–156.
19. Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
20. Gilinsky, A. S. (1951). Perceived size and distance in visual space. *Psychological Review, 58*, 460-482.
21. Gilinsky, A. S. (1955). The effect of attitude upon the perception of size. *Journal of Psychology, 68*, 173-192.
22. Goldstein, E. B. (1987). Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *Journal of Experimental Psychology, 13(2)*, 256–266.
23. Goldstein, E. B. (1988). Geometry or not geometry? Perceived orientation and spatial layout in pictures viewed at an angle. *Journal of Experimental Psychology: Human Perception and Performance, 14(2)*, 312–314.
24. Haber, R. N. (1985). Toward a theory of the perceived spatial layout of scenes. *Computer Vision, Graphics, and Image Processing, 31*, 282–321.
25. Hagen, M.A. (1976). Influence of picture surface and station point on the ability to compensate for oblique view in pictorial perception. *Developmental Psychology, 12*, 57–63.

26. Hagen, M. A. (1986). *Varieties of realism: Geometries of representational art*. Cambridge, England: Cambridge University Press.
27. Halloran, T. O. (1993). The frame turns also: Factors in differential rotation in pictures. *Perception & Psychophysics, 54(4)*, 496–508.
28. He, Z. J., & Ooi, T. L. (2000). Perceiving binocular depth with reference to a common surface. *Perception, 29*, 1313–1334.
29. He, Z. J., Wu, B., Ooi, T. L., Yarbrough, G., & Wu, J. (2004). Judging egocentric distance on the ground: Occlusion and surface integration. *Perception, 33*, 789–806.
30. Hochberg. J. (1962). The psychophysics of picture perception. *Audio-Visual Communication Review, 10*, 22-54.
31. Hochberg, J. (1978). Art and perception. In: *Handbook of Perception: Vol. X. Perceptual Ecology* (ed. Carterette, E. C. & Friedman, M. P.), 225–258. New York: Academic Press.
32. Kennedy, J. (1974). *A psychology of picture perception.* San Francisco: Jossey-Bass.
33. Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2004). Pointing out of the picture. *Perception, 33*, 513–530.
34. Koenderink, J. J., & van Doorn, A. J. (2008). The structure of visual spaces. *Journal of Mathematical Imaging and Vision, 31*, 171–187.
35. Koenderink, J. J., van Doorn, A. J. & Wagemans, J. (2011). Depth. *i-Perception, 2(6)*, 541–564.
36. Koenderink, J. J., & van Doorn, A. J. (2012). Gauge fields in pictorial space, *SIAM Journal on Imaging Sciences, 5(4)*, 1213–1233.
37. Kraft, R., & Green, J. S. (1989). Distance perception as a function of photographic area of view. *Perception & Psychophysics, 45*, 459-466.
38. Kubovy, M. (1986). *The psychology of perspective and renaissance art.* Cambridge: Cambridge University Press.
39. Madison, C., Thompson, W., Kersten, D., Shirley, P., & Smits, B. (2001). Use of inter-reflection and shadow for surface contact. *Perception & Psychophysics, 63*, 187-194.
40. Meng, J. C., & Sedgwick, H. A. (2001). Distance perception mediated through nested contact relations among surfaces. *Perception & Psychophysics, 63*, 1-15.
41. Meng, J. C., & Sedgwick, H. A. (2002). Distance perception across spatial discontinuities. *Perception & Psychophysics, 64*, 1-14.
42. Ni, R., Braunstein, M. L., & Andersen, G. J. (2004). Perception of scene layout from optical contact, shadows, and motion. *Perception, 33*, 1305-1318.
43. Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature, 414*, 197–200.
44. Ooi, T. L., Wu, B., & He, Z. J. (2006). Perceptual space in the dark affected by the intrinsic bias of the visual system. *Perception, 35*, 605–624.
45. Perkins, D. N. (1973). Compensating for distortion in viewing pictures obliquely. *Perception & Psychophysics, 14*, 13–18.
46. Perona, P. (2007). A new perspective on portraiture. *Journal of Vision, 7(9)*, 992-992.
47. Philbeck, J. W., & Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 72–85.
48. Pirenne, M. H. (1970). *Optics, Painting and Photography*. Cambridge, U.K.: Cambridge University Press.
49. Rogers, S. (1995). Perceiving pictorial space. In: *Perception of space and motion* (ed. Epstein, W. & Rogers, S.), 119–163. San Diego: Academic Press.
50. Rogers, B., & Gyani, A. (2010). Binocular disparities, motion parallax, and geometric perspective in Patrick Hughes's "reverspectives": Theoretical analysis and empirical findings. *Perception, 39*, 330–348.
51. Rosinski, R. R., & Farber, J. (1980). Compensation for viewing point in the perception of pictured space. In: *The Perception of Pictures* (ed. Hagen, M.A.), 137–176. New York: Academic.
52. Rosinski, R. R., Mulholland, T., Degelman, D., & Farber, J. (1980). Picture perception: An analysis of visual compensation. *Perception & Psychophysics, 28*, 521–526.
53. Sedgwick, H. (2003). Relating direct and indirect perception of spatial layout. In: *Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space* (ed. Hecht, H., Schwartz, R. & Atherton, M.), 215–238. Cambridge, MA: MIT Press.

54. Sinai, M. J., Ooi, T. L., & He, Z. J. (1998). Terrain influences the accurate judgment of distance. *Nature, 395*, 497-500.
55. Smith, O. W., & Gruber, H. (1958). Perception of depth in photographs. *Perceptual and Motor Skills, 8*, 307–313.
56. Todorovic, D. (2008). Is pictorial perception robust? The effect of the observer vantage point on the perceived depth structure of linear-perspective images. *Perception, 37*, 106–125.
57. Todorovic, D. (2009). The effect of observer vantage point on perceived distortions in linear perspective images. *Attention, Perception, & Psychophysics, 71(1)*, 183–193.
58. Toye, R. C. (1986). The effect of viewing position on the perceived layout of space. *Perception & Psychophysics, 40*, 85-92.
59. Vishwanath, D., Girshick, A. R., & Banks, M. S. (2005). Why pictures look right when viewed from the wrong place. *Nature Neuroscience, 8*, 1401–1410.
60. Wagemans, J., van Doorn, A. J., & Koenderink, J. J. (2011). Pictorial depth probed through relative sizes, *i-Perception, 2*, 992–1013.
61. Wagner, M. (2006). *The Geometries of Visual Space*. Mahwah, N.J.: Lawrence Erlbaum Associates.
62. Wallach, H., & Marshall, F. J. (1986). Shape constancy in pictorial representation. *Perception & Psychophysics, 39*, 233–235.
63. Wiest, W. M., & Bell, B. (1985). Stevens's exponent for psychophysical scaling of perceived, remembered and inferred distance. *Psychological Bulletin, 98*, 457-470.
64. Willats, J. (1997). *Art and representation*. Princeton, NJ: Princeton University Press.
65. Wu, B., Ooi, T. L., & He, Z. J. (2004). Perceiving distance accurately by a directional process of integrating ground information. *Nature, 428*, 73–77.
66. Yang, T., & Kubovy, M. (1999). Weakening the robustness of perspective: Evidence for a modified theory of compensation in picture perception. *Perception & Psychophysics, 61(3)*, 456–467.