

Depth Perception in Virtual Reality

Anja Matatko, Jürgen Bollmann, and Andreas Müller

Abstract Thematic 3D cityscapes represent a combination of the advantages of thematic 2D maps with those of 3D modelling. The application of thematic 3D cityscapes, however, poses the question of difficulties for orientation in the space represented. Here is where, among other things, factors of depth perception must be applied. This study by the University of Trier examines empirically the phenomenon of depth perception in thematic 3D cityscapes to be able to offer recommendations to makers of such models, recommendations for so-called depth cues which improve perception of the spatial environment and which can be technologically implemented. These recommendations are based on theoretical approaches to depth perception which were established in recent years in the field of perception psychology and geovisualisation.

1 Introduction

In recent years, development and use of geo-spatial 3D models has increased greatly. Of scientific interest are the processes of acquiring 3D geodata, concepts and methods of landscape or city modelling as well as system development for interactive and stereoscopic model presentation. In addition, research examines problems of acquiring and processing spatially based information and knowledge structures by model users. The great degree of optical immersion found with rendering 3D models leads to the fact that special questions of spatial or depth perception, such as factors for navigation and interactive processes, generally play only a subordinate role.

The effectivity of these factors, however, is of interest especially when structures, streets and other model elements are graphically linked to geo-spatial problems, which is the case in Virtual City Models. The term “Virtual City Model” covers thematic 2D maps, block models, photorealistic models and thematic 3D

A. Matatko (✉), J. Bollmann, and A. Müller
Cartography, University of Trier, Behringstrasse 54286, Trier, Germany
e-mail: matatko@uni-trier.de, bollmann@uni-trier.de, muellea@uni-trier.de

city models. The neglect of rendered elements in thematic 3D city models, such as object textures and colours leads to a reduction of possibilities for mental identification, orientation and recall. This situation can be seen in a thematic city model developed at the University of Trier, one in which rendered elements were greatly reduced: the model is described in Matatko et al. (2009, 2009). Figure 1 shows a detail of the model. Streets, houses and roofs represent classified thematic attributes. Such models can especially be used for urban development planning. In order to make mental processing of information in such models as effective as possible, however, our aim is to analyse the entire field of spatial and depth perception and study factors which can be integrated, if necessary, into thematic city models.

Some elements leading to depth perception are already implemented in the construction and programming of 3D models. The person using the 3D modelling and rendering software must also, if necessary, conceptualise and integrate additional elements. Thus, the question arises which elements facilitating the impression of spatial depth actually possess relevance in cartographically oriented 3D models. Our aim is to get information about the significance of each single depth cue. As most of the software automatically integrates several depth cues, we tried to isolate them as good as possible in order to get empirical evidences for their significance.

In this article, we initially discuss different concepts of depth perception and its components. Then, we focus on selected results of a comprehensive empirical study. A larger goal is to clarify relationships and degrees of effectivity as well as the possible applications of factors of depth perception when using virtual cityscapes.

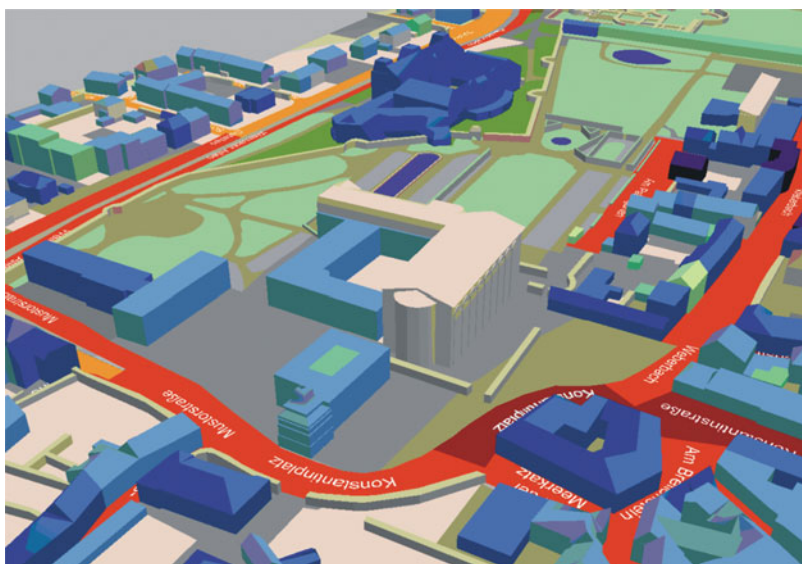


Fig. 1 Thematic 3D city model created by the University of Trier with representation of the living quality index as well as the public space quality index

2 Depth Perception

Examining spatial perception and its properties occupies an essential part of research into the psychology of human perception. Figure 2 represents the study topic in the psychology of perception according to Guski (1996), stating that all sense organs are relevant to acquiring information relevant to space. Not only visual perception but also the other sense organs lead to an impression of spatial depth, especially hearing. This study only treats questions of visual perception.

Goldstein (2002) differentiated between two basic theoretical approaches to explain depth perception: the cue theory and the ecological approach proposed by J. J. Gibson (1973), with the cue theory following the constructivistic approach. It assumes that the observer plays an active role in the perception process. The cue theory intends to identify information on the retinal image, processed information matching the depth of the actual, existing world. The ecological approach emphasises the observer and his/her interaction with the environment.

This chapter is based on different schemata of depth cues, as proposed by Goldstein (2002), Albertz (1997), and Ware (2004). As there is no generally accepted schema about depth cues and their impact on depth perception (and our aim is not to create one), we summarize different depth cues and depth perception gradients from different authors.

2.1 The Basic Approach of the Theory of Depth Cues

The theory of depth cues and related unconscious information processing was proposed by Helmholtz as early as 1896 (von Helmholtz 1896). According to Albertz (1997), the third dimension within the visual perception process is “constructed”

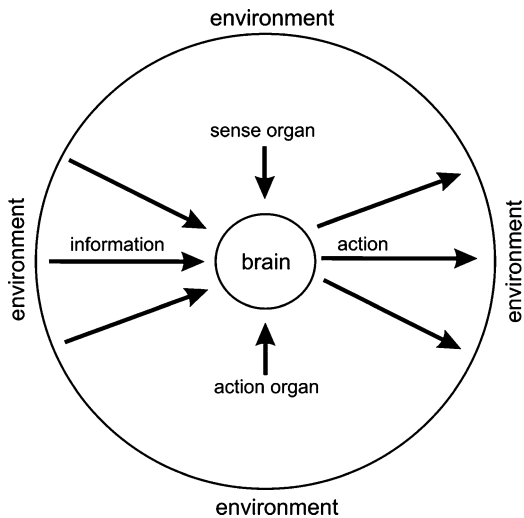


Fig. 2 Study topics in the psychology of perception (Altered on the basis of Guski 1996)

from the signals arriving on the retina. This process requires several elements of spatial perception to achieve depth perception. The impression of depth arises through combining experience, environmental stimuli and images of these stimuli on the retina.

These elements, called depth cues, may, as proposed by Goldstein (2002), be divided into four groups: oculomotor cues, pictorial cues, motion-produced cues and binocular disparity. Other groupings differentiate between binocular and monocular cues. In Goldstein, the first three groups comprise monocular cues, that is, they can also be processed with only one eye.

2.1.1 Oculomotor Cues

In contrast to other depth cues, oculomotor cues are perceptible because of the alterations in the eye during movement of the eye as it perceives space. In fixating on a point in space, the lens of the human eye becomes curved and accommodates. The further the eye moves its focus from the fixation point, the more blurred the image becomes. Spatial depth can thus be estimated. However, the observer cannot recognise whether objects lie in front of or behind the fixation point. With objects far removed from the fixation point, the effect of the blur becomes weaker for optical reasons. Goldstein (2002) numbers convergence and accommodation among the oculomotor cues. He defines convergence as the movement of muscles causing an inward movement of the eyes. He defines accommodation as the bulging of the lens while focusing on near objects. Kelle (1995) considers accommodation subordinate, as he views the reaction of the eye muscles as one based on an already realised depth perception.

2.1.2 Pictorial Cues

Pictorial cues comprise the largest share of monocular cues. They are based on interpretations of information derived from images. As the retina itself receives only a two-dimensional copy of reality, the observer must use optical links to obtain a three-dimensional impression. Six pictorial cues are introduced in the following:

- *Overlap*. A topological alignment of depth can be created when objects further in the background are occluded by those closer to the observer.
- *Size in the field of view*. The greater the portion of the object's surface is in the field of vision, the larger the object will be estimated to be.
- *Height in the field of view*. Objects located higher up in the field of vision are perceived to be further away.
- *Atmospheric perspective*. Contrast reduction with very great distances, owing to particles in the air ("fogging effect").
- *Familiar size*. The observer knows from experience that objects become smaller the further away they are. If either the object size or the distance is known, the observer can draw corresponding conclusions concerning the unknown parameter.

- *Linear perspective*. Parallel lines in space converge toward a common vanishing point, closely related to texture gradient.

2.1.3 Motion-Produced Cues

Besides depth cues available to a static observer, other depth cues exist which become effective only when the observer moves.

- *Motion parallax*. Near objects appear blurred to the observer and move quickly, whereas objects further away move more slowly and are more easily identified. Albertz (1997) accorded motion parallax “enormous importance for perceiving our environment”.
- *Deletion and accretion*. In non-perpendicular movements, objects not at the same distance tend to appear to move relative to one another. This effect is called deletion when the object in the background becomes increasingly occluded by the object in front of it. It is called accretion when the object in the rear emerges from behind the object in front.

2.1.4 Binocular Disparity

Binocular disparity is the most important cue for depth. It is defined as the acquisition of a spatial impression through the effect of binocular disparity, an effect arising from the off-set position of the two human eyes. Due to the different positions, images are created on both retinas from two different perspectives. In 1959, Béla Julesz was able to confirm binocular disparity with the help of his Random Dot Stereogrammes, in which a spatial impression could be created without using further depth cues.

2.1.5 Effectivity of Depth Cues

The individual depth cues develop their effectivity at different distances. For example, atmospheric perspective is effective only at distances greater than 30 m (100 ft), whereas the effect of binocular disparity is already greatly reduced after a few metres, and, according to Buchroithner and Schenkel (2001) ceases completely at 10 m. Further effects are represented in Table 1.

2.2 Depth Perception Gradients

According to Gibson (1973), the environment consists of stimuli which can be received and processed by living things according to their abilities. The author lists several requirements for research into depth perception: First, depth perception

Table 1 Effectivity of selected depth cues (Altered on the basis of Goldstein 2002)

Depth cue	<2 m	2–30 m	>30 m
Overlap	Yes	Yes	No
Familiar size	Yes	Yes	Yes
Oculomotor cues	Yes	No	No
Binocular disparity	Yes	No	No
Atmospheric perspective	No	No	Yes

should be examined by a moving observer, as dynamics prevail in reality. Second, in contrast to the theory of depth cues, it is not the image on the retina to be analysed but the information available to the optic array. Gibson (1973) defines various elements of optic arrays similar to depth cues, elements leading to depth perception, among them texture gradients.

Albertz (1997) expanded and differentiated the effect and the significance of texture gradients, which are an optical transitional phenomenon continually extending into spatial depth. The author thus differentiates *size*, *density* and *form* gradients, all of which are interrelated to the conditions of the central perspective structure of space. For example, windows leading off to the vanishing point appear to be continually reduced in their size and intervals as well as to systematically be altered in form, thus creating a gradual but continuous gradation. Closely related to this phenomenon is the actual texture gradient, in which, with increasing spatial depth, the grain of the surface pattern, for example, a cobbled street or vegetation, becomes optically finer.

Additional gradients are the *contrast* and *colour* gradients arising from atmospheric phenomena as characteristics of the so-called atmospheric perspective. According to Albertz (1997), continuous brightness gradients created by the formation of cast shadows on objects have great significance. Sequences of projected views of objects extending into spatial depth can result through increasing gradual distortion and deformation through cast shadows and reflections.

2.3 *Status of Current Research in Cartography and Geovisualisation*

Thus far, hardly any concrete and complete empirical studies on depth perception have been published in the fields of cartography or geovisualisation. Especially informative for this study, however, are the publications of Albertz (1997) and Buchroithner (2001). The latter work differentiates various process groups creating spatial images which differ in the depth cues used, resulting in degrees of perception and immersion. The authors stipulate that for variable visualisation tasks of geodata the individual processes must be evaluated on the basis of test applications. They introduce a selection of processes based on estimates and point to continuing studies still required for evaluation of this pre-selection.

As mentioned above, Albertz (1997) expressly differentiates among the effects of various perception gradients and places them in the context of the active observer in the perception space. Thus taken into consideration, on the one hand, are the properties of optical processes into spatial depth known from the observer’s environment; and addressed, on the other hand, are the active positioning as well as mental attitude and direction of the observer. Both aspects are of special interest for the user-oriented, active mastering of virtual space. Kraak (1988), Meng (2002), Jobst and Germanchis (2007), for example, all require an ample number of depth cues as well as sufficient a priori knowledge for the observer to be able to acquire a spatial impression.

On the whole, an acceptable theoretical basis resulted in being able to develop the tasks in the following study and to classify the empirical conclusions.

3 Empirical Analysis of Selected Depth Cues

3.1 Methodical Background

The empirical study was conducted with the aid of the 3D model of the Trier inner city created by the cartography department. In ten studies, eight depth cues were examined, represented in Table 2. Along with the above-named theoretically documented factors, additional influential factors were examined experimentally, factors from the field of spatial representation and perception. A prerequisite to studying a depth cue is that it both had to be implementable with the existing technological capabilities and had to be regarded as relevant also for thematic 3D models. The software used for modelling and navigation in the 3D models was a combination of 3D Studio Max and Virtools. The latter facilitates creating application-based programming of navigation and interaction functions based on a modular system. The main problem in the studies was that the individual depth cues were difficult to isolate from one another, as several depth cues normally exist automatically. This was especially true of the study of overlap. Each study had a sample size between 25 (for those using eye tracking) and 200 (for those using online surveys).

Table 2 Studies of depth perception in 3D city models

Depth cues	Implementation
Binocular disparity	Time consuming
Linear perspective	Automatic
Motion parallax	Automatic with dynamics (movement)
Overlap	Automatic
Texture gradient	Automatic
Contrast gradient	Integrable
Depth of focus	Complex
Viewing angle	Automatic

Table 3 Participant characteristics of selected studies of depth perception

Study	Texture gradient	Depth of field	Contrast gradient	Motion parallax
Number of participants	Online survey: 200 Eye tracking: 20 students	Online survey: 400 students	Eye tracking: 40 students	Eye tracking: 28 students
Intellectual background pre-experience	n.a.	Non-experienced and experienced computer users	n.a.	Non-experienced and experienced VR-users

Table 2 represents the software implementation capabilities of all the examined depth cues.

In the following sections, four studies will be presented in detail: texture gradient, depth of field, contrast gradient and motion parallax. Table 3 compares the exact number of study participants, their intellectual background and pre-experience.

3.2 Texture Gradient

In their study, Bott et al. (2007) have documented the effect of texture gradients empirically. Each task contains distance estimations. The distances are to be given as relative or absolute but with an indicated reference distance. The ground surface for the study is represented in three different scenes: without texture, without texture gradient (distorted texture with consistent grain; that is, the intervals toward the rear do not become smaller), with texture gradient. To avoid perspective influences, neither an unstructured asphalt surface nor a paved structure with regularly ordered stone was used.

The results of the study are displayed in Fig. 3. The largest percentage of correct answers (73%) occurred with scenes showing undistorted texture (that is, with natural texture gradient). The response ratio was about identical for the scenes using the distorted texture or no texture, with approximately 30% accurate estimations.

With the aid of eye tracking, it could be shown that the test persons oriented themselves on the ground texture in the landscape to estimate the distances. “Hot spot” images were used to illustrate the eye movements. Figure 4 contrasts the results of a scene with and without ground texture. In the scene without texture, it can clearly be seen that the test persons sought other orientation points in the landscape to be able to answer the question. In the scene with ground texture, the fixations are concentrated on the central area of the surrounding ground texture. The behaviour of test person shown in Fig. 4 is representative for all test persons participating in the study of texture gradient.

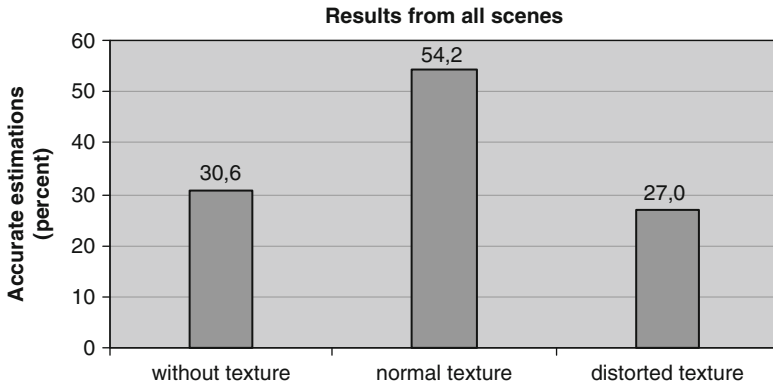


Fig. 3 Results from the test images with texture, without texture and with distorted texture (Altered on the basis of Bott et al. 2007)

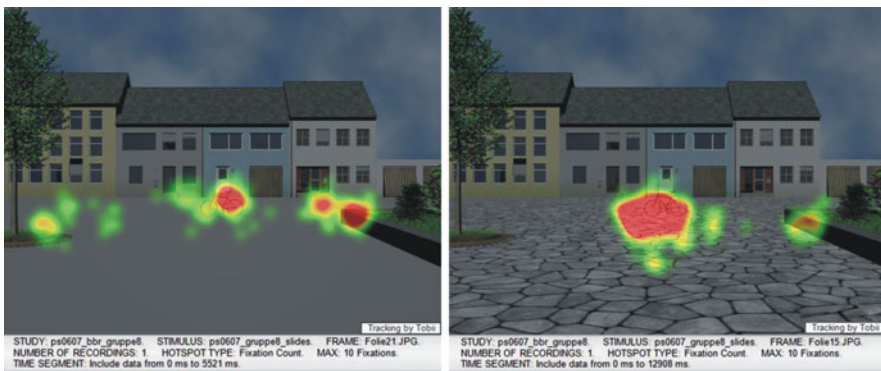


Fig. 4 Hot spot analysis of scenes without (left) and with (right) texture (Altered on the basis of Bott et al. 2007)

3.3 Depth of Field

Depth of field is a phenomenon that plays a role primarily in photography and video recording techniques, in which an emphasis is placed on a desired section of an image by setting the lens for sharp or unsharp areas. Greiling and Marx (2007) studied the effect of depth of field created through optical structuring of image scenes in city models as a possible factor in depth perception. Test persons had to solve tasks within several scenes with different depth of field settings. Figure 5 shows two details of the study. As part of an Internet survey, settings and acceptance criteria were ascertained in addition to the experiment.

The results of the study were surprising in that, with depth of field, subjective impression and objective action were widely divergent. In the survey, 90% of the test persons declared that they preferred scenes with sharp depth of field.

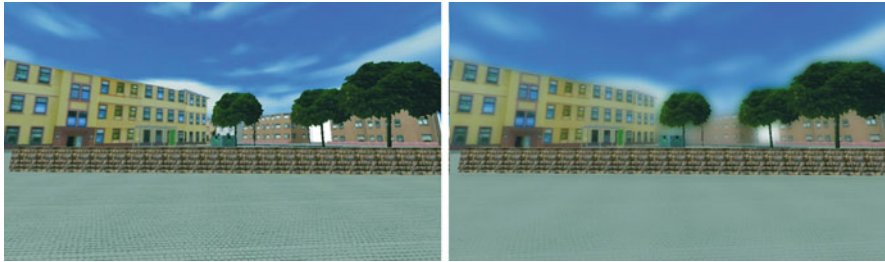


Fig. 5 Landscape completely sharp (*left*) and with foreground focus (*right*) (Altered on the basis of Greiling and Marx 2007)

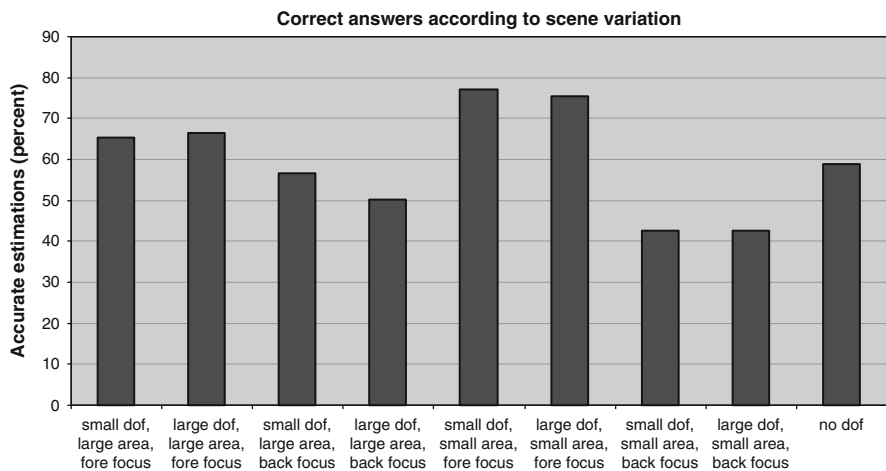


Fig. 6 Success ratio with variations in depth of field, size of area and kind of focus (dof = depth of field) (Altered on the basis of Greiling and Marx 2007)

Figure 6 shows, however, that the success rate for scenes with focus in the foreground rises and thus probably supports the solution to the tasks through effects of depth of field. There are few differences between men and women. There is also no significant difference in the comparison between larger and smaller areas.

3.4 Contrast Gradient

Voshaar and Metzger (2007) have represented the effect of contrast gradients. Combined with colour gradient, it forms the atmospheric perspective. For technical reasons, only the contrast gradient was studied, as the colour gradient is not implemented in Virtools. During the study, the fixation movements of the test persons were recorded and evaluated using the Tobii Eye Tracker. Details from

the city model were shown at different positions. The tasks related to estimating distances of objects to one another, that is, test persons were asked to estimate relative values instead of absolute distances. Figure 7 represents two scenes from the study: without decreased contrast on the left and with decreased contrast on the right due to haze.

The evaluation proved to be a problem because, in general, the estimations were grossly incorrect, regardless of the amount of contrast decrease. Figure 8 shows that, with a great amount of haze, the percentage of correct estimations clearly increased (45%) in comparison to the scenes without or with little haze.

Information about spatial orientation can also be obtained from the parameters of eye movement. In the scenes without decreased contrast due to haze, more fixations were counted, ones which were shorter, on the average, than in the scenes with decreased contrast. With the example of one scene, Fig. 9 represents the average fixation time. Based on the shorter fixation time, we can conclude that the test persons were better able to orient themselves in the scene with decreased



Fig. 7 Landscape without (left) and with (right) decreasing contrast due to haze (Altered on the basis of Voshaar and Metzger 2007)

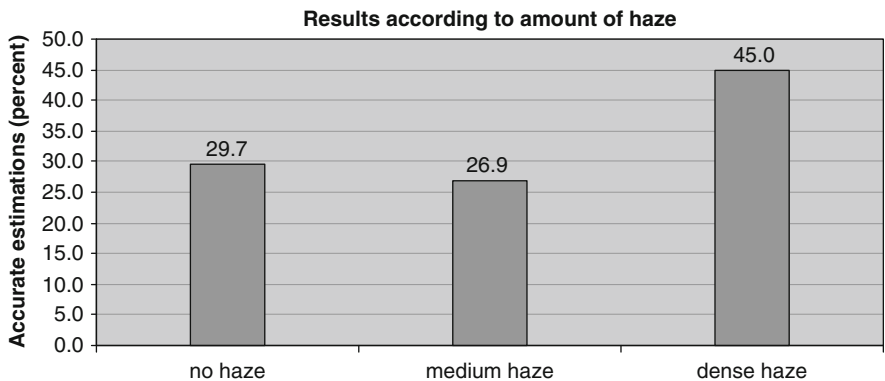


Fig. 8 Evaluation of the results of estimations according to amount of haze (Altered on the basis of Voshaar and Metzger 2007)

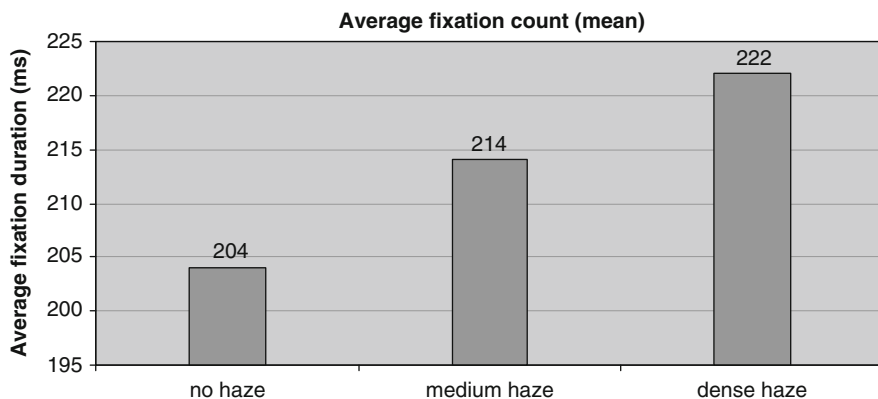


Fig. 9 Average fixation time with scenes with greatly different contrast effect (Altered on the basis of Voshaar and Metzger 2007)

contrast. The eye movements occur less aimlessly. Objects were fixated longer, indicating that more exact information gathering was possible.

3.5 Motion Parallax

To find out whether motion parallax increased depth perception, the test persons were to estimate the distances of objects. Glass recycling bins were selected as actual objects from the cityscape; the effects of other cues such as cast shadows or overlaps were limited by configuring the parameters in Virtools. Scenes with and without movement were juxtaposed. Each movement occurred parallel to the objects, at two different speeds.

The results determined that no significant difference exists between the slow and the fast camera movement. However, in respect to the number of correct estimations, distinctly appreciable differences exist between the static and the dynamic image. The mean percentage of correctly solved tasks lay higher with slower or faster camera movement than for the static view. In addition, the scatter was less, especially with the slow camera movement, than with the use of screenshots. It can thus be concluded that motion parallax contributes essentially to depth perception in the thematic landscape.

3.6 Further Study Results

3.6.1 Binocular Disparity

In the study of binocular disparity, difficulties occurred because this disparity, as described above, is effective only at close range. Thus the question arises whether

stereoscopy is of great significance at all for perception in a thematic 3D cityscape. This applies primarily to mental processes in which acquiring a spatial overview and the mental construction of spatial coherence is emphasised.

3.6.2 Linear Perspective

If central perspective is exchanged for parallel perspective, the exchange has a negative effect on achieving perception. Sex-related differences exist insofar as the answers of men were more correct than those of women. Similarly, positive correlations exist between the prior experience with 3D models and the correctness of answers. With a dynamic representation, the test persons were able to alter the viewing position in such a way that they could select between parallel perspective and central perspective. The test persons preferred the central perspective.

3.6.3 Overlap

In the study of overlap, both the number of objects to be identified and the number of overlapped objects was varied. Fixations were recorded using eye tracking. It was shown that, as the number of objects in a landscape to be identified increased, the response time rose. No significant results were achieved in respect to variation in the number of the overlapped objects, a result associated with the fact that, among other things, the effect of overlap can be substantiated only with difficulty, as this depth cue can hardly be observed isolated from others. The visual use of overlapping objects could, however, be documented with the aid of hot-spot analyses of the fixations.

3.6.4 Viewing Angle

In 3D models, the viewing angle can be varied at liberty. Thus the question arises which viewing angle can be considered ideal. This applies especially to models where no completely free navigation exists but where, for example, aerial tours are conducted along a defined path. The study was able to document that an angle of approximately 15° was considered ideal.

4 Conclusion

This study documents the relevance of individual depth cues in spatial perception in thematic 3D models. The isolation of individual depth cues, however, proved to be a problem for the study, for example, with overlap and linear perspective, as these occur of necessity in 3D models where it is difficult to limit them graphically as

well as in their effect. In general, it was attempted to isolate the studied depth cues as far as possible.

A further step in the study was to conduct a classification of depth cues according to their relative importance to assist the makers of 3D cityscapes in creating better user-oriented models, so that such models can simplify orientation and mental information processing by the user.

It was shown that contrast gradient, motion parallax and texture gradient are especially effective for the concrete application of the Trier 3D city model. However, the results must be further examined, in particular with the increase of free navigation in the 3D models, and adapted to new technological possibilities and their ensuing conditions of perception in models.

References

- Albertz, J.: Die dritte Dimension – Elemente der räumlichen Wahrnehmung. In: Albertz, J. (ed.): *Wahrnehmung und Wirklichkeit*. Schriftenreihe der Freien Akademie, vol. 17, pp. 81–108. Freie Akademie, Berlin (1997)
- Bott, F., Mertens, N., Lutz, R.: *Empirische Untersuchung zur Wirkung des Texturgradienten anhand einer 3D-modellierten Szene aus dem Stadtteil Trier-Nord*. Unpublished project report, University of Trier, Trier (2007)
- Buchroithner, M.F., Schenkel, R.: 3D-Visualisierung von Geodaten – perceptionstheoretische und präsentationstechnische Grundlagen. In: Koch, W.G. (ed.): *Theorie 2000*. Vorträge des kartographischen Symposiums am 17/18 November 2000 an der TU Dresden. Kartographische Bausteine, vol. 19, pp. 113–120. Institut für Kartographie, TU Dresden, Dresden (2001)
- Gibson, J.J.: *Die Sinne und der Prozeß der Wahrnehmung*. Verlag Hans Huber, Bern, Stuttgart, Vienna (1973) [The English original is titled: *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston (1966)]
- Goldstein, E.B.: *Sensation and Perception*. 6th ed., Wadsworth Group, Pacific Grove (2002)
- Greiling, M., Marx, A.: *Projektbericht zur Untersuchung des Tiefenfaktors ‘Tiefen(un)schärfe’ auf seine Relevanz bei der Tiefenwahrnehmung des Menschen anhand eines Online-Experiments*. Unpublished project report, University of Trier, Trier (2007)
- Guski, R.: *Wahrnehmen. Eine Einführung in die Psychologie der menschlichen Informationsaufnahme*. Kohlhammer, Stuttgart (1996)
- Jobst, M., Germanchis, T.: *The Employment of 3D in Cartography – An Overview*. In: Cartwright, W., Peterson, M.P., Gartner, G. (eds.): *Multimedia Cartography*, 2nd ed., pp. 217–228. Springer, Berlin Heidelberg (2007)
- Kelle, O.: *Dynamische Tiefenwahrnehmungskriterien in Computergenerierten Interaktiven Szene und Virtuellen Simulationsumgebungen*. VDI-Verlag GmbH, Düsseldorf (1995)
- Kraak, M.-J.: *Computer-Assisted Cartographical Three-Dimensional Imaging Techniques*. Delft University Press, Delft (1988)
- Matakto, A., Müller, A., Bollmann, J.: *Ein Wohnumfeldqualitätsindex für Trier*. In: Strobl, J., Blaschke, T., Griesebner, G. (eds.): *Angewandte Geoinformatik 2009*, pp. 826–835. Wichmann, Heidelberg (2009)
- Matakto, A., Müller, A., Bollmann, J.: *Thematische 3D-Stadtmodelle: Modellierung, Anwendung und empirische Befunde*. In: *Mitteilungen aus dem BKG, Band 45: Arbeitsgruppe Automation in der Kartographie – Tagung (2009)*, S99–S106
- Meng, L.: *How can 3D Geovisualization Please Users Eyes Better?* *Geoinformatics Magazine for Geo-IT Professionals*, vol. 5, 34–35 (2002)

- von Helmholtz, H.: Handbuch der Physiologischen Optik. 2nd ed., Verlag von Leopold Voss, Hamburg (1896)
- Voshaar, E., Metzger, M.: Bericht zur kartographischen Projektstudie. Unpublished project report, University of Trier, Trier (2007)
- Ware, C.: Information Visualization. Perception for Design. 2nd ed., Morgan Kaufmann, San Francisco (2004)