Towards Attention-Centered Interfaces: An Aesthetic Evaluation of Perspective with Eye Tracking

INGMAR S. FRANKE, SEBASTIAN PANNASCH, JENS R. HELMERT, ROBERT RIEGER, RAINER GROH, and BORIS M. VELICHKOVSKY
Technische Universität Dresden

The established method of representing three-dimensional space on a two-dimensional surface involves camera based, point of regard systems, comparable in design to the early "camera obscura". However, geometrical limitations of such models lead to distortions of perspective when projected. This research investigated the influence of single- versus multi-perspectives on aesthetic choices within one image. A clear perceptual bias towards multi-perspective images was found, additionally supported by an eye tracking study. We propose that human users are more attracted by multi-perspective images, which emphasise the "semantic foci" of the scene, than by those being synthesized statically with only one geometrical prospect.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—*Human information processing*; H.5.2 [**Information Interfaces and Presentation (e.g., HCL)**]: User Interfaces—*Evaluation / methodology, user-centred design*; J.4 [**Social and Behavioral Sciences**]: —*Psychology*

General Terms: Design

Additional Key Words and Phrases: Eye tracking, subjective evaluation, perspective projection, scene perception

ACM Reference Format:

Franke, I. S., Pannasch, S., Helmert, J. R., Rieger, R., Groh, R., and Velichkovsky, B. M. 2008. Towards attention-centred interfaces: An aesthetic evaluation of perspective with eye tracking. ACM Trans. Multimedia Comput. Commun. Appl. 4, 3, Article 18 (August 2008), 13 pages. DOI = 10.1145/1386109.1386111 http://doi.acm.org/10.1145/1386109.1386111

1. INTRODUCTION

Human vision is designed to infer three-dimensional information from two-dimensional images. Indeed, the image that forms on the retina is two-dimensional. There are many means by which three dimensions can be perceived, through variations of light and shade, relative size and dimensions of objects in perspective and the fact that we have two eyes, and hence two separate points of view on the three-dimensional world. When we talk about graphical representation of space in everyday life, we are generally referring to photographs, computer graphics, and print media, which depict a "virtual"

This research was supported by two EU FP6 NEST Pathfinder grants, PERCEPT 043261 and MINET 043297 to Sebastian Pannach, Jens R. Helmert, and Boris M. Velichkovsky.

Authors' addresses: I. S. Franke, R. Rieger, and R. Groh, Department of Computer Science, Institute of Software and Multimedia-Technology, Technische Universität Dresden, 01187 Dresden; email: {if4, robert.rieger, rg5}@mail.inf.tu-dresden.de; S. Pannasch, J. R. Helmert, B. M. Velichkovsky, Applied Cognitive Research Unit, Institute of Psychology III, Technische Universität Dresden, 01069 Dresden; email: {pannasch, helmert, velich}@psychomail.tu-dresden.de.

three-dimensional space on an actual two-dimensional surface image. Insofar as the observer can be seen as the user of the image, the image itself is the interface between the observer and the inferred space.

In terms of usability, this interface should allow the observer to perceive spatial information as effectively and efficiently as possible [Jokela et al. 2003]. An objective and simple way of projecting objects on an image plane is the *camera obscura* method. Using projected perspective, it is the basis for the standard camera model of computer graphics as well as photography. Yet, its limitations become obvious as soon as objects fall outside the central line of projection. The dimensions of the projected objects change significantly with increasing digression from its central line, so that sphere at the periphery of the projected image would be depicted elliptically.

The Old Masters of the Renaissance also used the camera obscura model to geometrically construct various vistas on canvas [Edgerton 1975]. They developed a method that prevented the distorted impressions associated with projected perspective. This technique was principally employed by veduta painters (Italian: veduta "view") in order to produce realistic and highly detailed cityscape or landscape paintings. They counterbalanced the distortions associated with having an entire scene projected through a single vanishing point by using a sophisticated combination of several perspectives, resulting in a perceptibly more realistic work [Groh 2005]. This represents the first attempt to construct two-dimensional images which is a medium of active interaction between the human observer and the image itself.

The present studies are based on two approaches to image production: the current point of regard approach used in computer graphics and the historical approach of veduta painting. Both aim at generating realistic, highly detailed representations of three-dimensional space; veduta paintings are highly detailed and large scale, and computer graphics aim at increasingly detailed rendering of realistic virtual environments. In this regard, the term "realistic" should be defined as matching *the human experience of reality* as closely as possible. Hence, the subsequent section will discuss these approaches to the design of "virtual realities" in terms of the human experience.

2. APPROACHES

This work is concerned with methods of representing three-dimensional space on two dimensions in order to create a maximally realistic experience for the viewer. Although certain differences between current and historical approaches are immediately apparent, the motivation for this research was to compare the efficacy of these approaches systematically and analyze their relative efficiency based on the aesthetic choices and ongoing activities that represent the users' subjective perceptions of the space.

2.1 Human Vision

Comparable to the camera obscura, the human eye consists of a refractive and light gathering medium (the lens and eyeball) and a photosensitive system (retina). However, the main difference to the camera obscura is the heterogeneous sensitivity of the retina to light and color. Although normal vision extends to about 100 degrees of vision [Findlay and Gilchrist 2003], only the central two degrees are registered with color and in high resolution due to the physical properties of the retina [von Helmholtz 1867; Palmer 1999]. The center of the retina, the *fovea centralis*, contains tightly packed photoreceptors sensitive to color and individually connected to the optic nerve. Outside this central region, light falling on the retina is less differentiated, and its color is not registered. Hence, visual acuity, as we experience it as a rich and detailed three-dimensional space, is formed based on the eye moving through various viewpoints and collecting detailed information from only a very narrow field of view each time it stops. Each time the eye moves the retinal image is shifted, in much the same way as the image recorded by a camera shifts if the camera is quickly turned or moved. Due to a suppression of visual information processing during such a shift, one does not notice this potentially disconcerting registration of information (first described by Dodge [1900]). From the brief stops the eye makes while scanning the visual world, and the resulting low-grade images registered by the retina, the visual cortex constructs a rich, detailed, colorful and spatially reasonable world in which we can manoeuvre ourselves and the objects around us with ease and according to our current goals. Thus, for a human observer, each visual space is constructed from many perspectives depending on the distribution of attention and resulting changes in gaze direction over time. The reality perceived by a single human user is a result of combining multi-perspectives into one, perceived three-dimensional space. In contrast to human visual perception, computer-generated "virtual" spaces usually result from one vanishing point or point of regard per image.

2.2 Perspective Projection in Computer Graphics

This section describes how three-dimensional space is transferred into a two-dimen-sional image in computer generated graphics. Object space can be represented based on a three-dimensional model containing polygonal objects or any other three-dimensional set of data. The image space is a two-dimensional image displayed onscreen. In order for the viewer to perceive three-dimensional objects, computer graphics makes use of projections, that is, transformations from higher dimensional into lower dimensional mathematical spaces. Most visualization systems use perspective projection [Rauber 1993; Foley 1999; Watt 1993]. As in the camera obscura model, perspective is geometrically determined by a central line of projection, which is transferred onto two dimensions. The image results from the intersection points of the projection plane with all connecting vectors from the centre of projection to the three-dimensional points. The camera model of computer graphics defines the centre of projection, the viewing direction, the field of view and the aspect ratio of the projection plane as well as a near and a far clipping plane. Thus, it implements the perspective projection [Angel 1997]. Furthermore, these parameters define a viewing frustum, that is, the part of the object space that is being rendered. Transformation of object space coordinates into screen space coordinates is specified by the rendering pipeline [Watt 1993]. In doing so, the following steps are executed consecutively: model view transformation, clipping, projection transformation and finally rasterization for displaying the image on the screen. In this article, the model view transformation is relevant since it produces a perspective shortening along the viewing direction of the camera. Hence, objects will be diminished in size the further they are positioned from the center of the projection.

Although this transformation takes place in order to imitate human vision, object space angles and distances that are not parallel to the projection plane have different values in image space. Also, parallel object space edges only run parallel in image space if they are parallel to the projection plane; otherwise, these projected edges converge in multiple vanishing points. If object space edges run orthogonal to the projection plane, this point is called the *principal vanishing point*. The principal vanishing point of an image is defined as the perpendicular base point of the camera's viewing vector at the projection plane. Furthermore, it is the intersection point between horizon and sagittal line within the image. By default, this point is located in the center of the image due to the symmetric definition of the viewing frustum [Angel 1997; Franke et al. 2005]. In other words, an image created via perspective projection has exactly one principal vanishing point. However, the perspective shortening also leads to an elliptic depiction of eccentrically positioned spherical objects. [Groh 2005]. This can be clearly seen in the church cupola (see Figure 1). The limitation of the standard camera model is that it defines one single mathematical center of projection, and therefore only one principal vanishing point. The distortion of spherical objects is a primary disturbance. There are several approaches within computer graphics towards resolving the distortion effects emerging from perspective projection. For example, extending the camera model [Agrawala et al. 2000] through a multi-projection rendering algorithm using multiple local

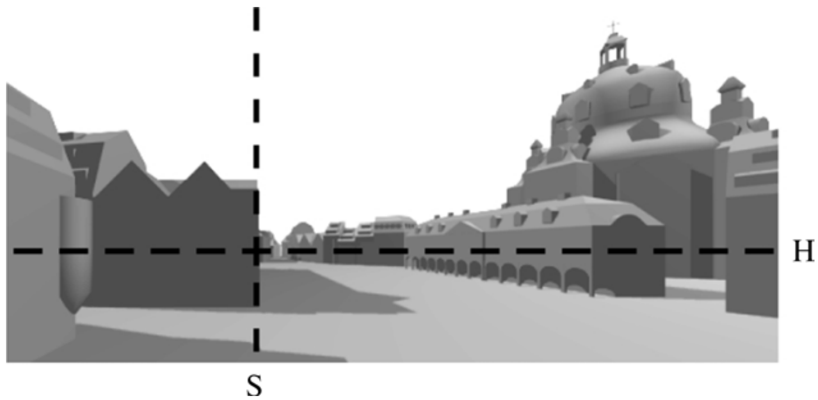

Fig. 1. Image of a modeled 3D scene. The principal vanishing point (S, H) of the image results from the vanishing point of all object space edges that are orthogonal to the projection plane. The ellipsoid form of the church cupola is evident.

cameras and manipulating z-buffer calculations. Other approaches generally concern the generation of multi-perspective images using image based rendering [Yu and Mcmillan 2004; Rademacher and Bishop 1998] or data from video material [Zomet et al. 2003; Roman et al. 2004]. Yet, all approaches neglect to address the constant fusion of multiple perspectives in normal human vision. The motivation for our work lies in applying these approaches to creating a realistic experience of three-dimensional space from a two-dimensional projection, using what is understood about human visual perception and how normal vision combines multiple perspectives. During the Renaissance, veduta painters were aware of the problem of distortion from single viewpoints, and successfully circumevented it, as the next section will show.

2.3 Veduta Painting

Veduta painters also used mulitple perspectives [Klotz 1990]. To record realistic images of the world on canvas, they operated to integrate these perspectives. Several viewpoints were integrated in the geometrical construction of the image, and combined to form one two-dimensional image [Groh 2005]. At this time, the camera obscura was used like an eye and was systematically oriented to certain objects [Groh et al. 2006]. Thus, each focused object received its own image with its own principal vanishing point that was subsequently integrated into the overall image. Their modus operandi reflects the attentive directing of gaze around a three-dimensional space basic to human vision and perception [Velichkovsky and Hansen 1996; Palmer 1999]. Hence, these images shall be referred to as *perceptually realistic*. Figure 2 shows a painting by Canaletto and its covert principal vanishing points. Within this image there are two principal vanishing points: the intersection point (S1, H) of sagittal line S1 and horizon H marking the first general point of regard, that is, from the main perspective of the image. The second principal vanishing point (S2, H) results from the special treatment of the *Church of Our Lady* [Groh 2000]. There are obvious differences between Figure 2 and Figure 1 concerning perspective distortions, but it is not immediately clear how these differences might influence subjective visual perception and its related aesthetic decisions and functionalities.

3. EXPERIMENTS

The aim of these experiments was to assess the effect of multi- versus single-perspective images by means of the users' aesthetic judgments and eye movement behaviors. In Experiment 1, subjects were

Fig. 2. Canaletto: “The new market in Dresden of Moritzstrasse”, 1750. The main perspective is indicated by intersection point (S1,H). Additionally the Church of Our Lady was treated with an additional vanishing point (S2,H). Reprinted with permission by the Dresden State Art Collections, Germany.

asked to make a decision between a single- and a multi-perspective version of the same image. In Experiment 2, we were interested in whether these judgments were related to the pattern of eye movements made by subjects during viewing. We distinguished between three categories of images: computer-generated images, photographs and veduta paintings. In the first two categories, images are usually created using single-perspective projection, whereas the latter consists of multi-perspectives. To compare the relative effect of multi-perspective versus single-perspective images, comparison pairs were first produced. Figure 3 illustrates the methods used to create image pairs. In generating multi-perspective images from computer-generated single-perspective ones, we applied a technique of correcting the proportions of objects that were subject to perspective projection distortions [Franke et al. 2007]. This technique is similar to that used by Agrawala et al. [2000]. In generating multi-perspective images from photographs, we used the approach presented by Zorin and Barr [1995]. Generating single-perspective images from paintings was more involved, as it was first necessary to identify the original vanishing points. Using manual-image based methods mentioned in Groh [2005], we analyzed the image for vanishing edges and points. Applying this information, we were able to restructure undistorted objects into the overall single-perspective geometry.

3.1 EXPERIMENT 1

Participants. In Experiment 1, 451 participants (306 male, 145 female) aged 17 to 39 (mean 20.7 years, SD 2.1 years) took part. All subjects were students of the Technische Universität Dresden. About 80 percent of the subjects were students of computer science in their first year of their studies. The remaining 20 percent were students from other departments, such as psychology, electrical engineering, industrial engineering and transportation science. All subjects were naive with respect to the purpose of the study.

Stimuli. Seventeen stimuli were presented. Each stimulus consisted of a pair of images, arranged side by side. Images differed in size, ranging from 460×290 to 468×719 pixels. Seven images

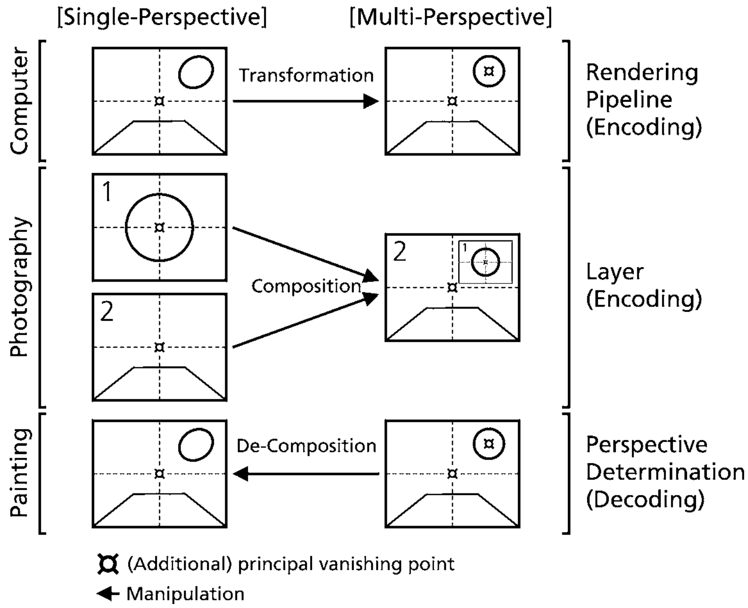

Fig. 3. Schema visualising the generation from single- to multi-perspective images and vice-versa.

Fig. 4. Examples of the stimuli of all categories: Computer rendered scene (upper left), photograph (lower left) and painting (right). Each pair consists of a single-perspective (left) and a multi-perspective (right) image version. "The Geographer" by Vermeer van Delft is reprinted with permission by the Städel Museum, Frankfurt am Main, Germany.

contained pairs of paintings, four were pairs of real photographs and three depicted pairs of computer rendered three-dimensional scenarios (see Figure 4 for examples of each category). One image of each pair followed the rules of perspective projection, while the other was generated according to the multi-perspective approach. Catch trials containing exactly the same control images (one painting, one three-dimensional scene and one photograph) were also presented in order to control possible learned response tendencies.

Procedure. After a short introduction at the beginning of a regular lecture the questionnaires were delivered to the participants in the lecture hall (35.5 m × 12 m). In order to control for possible influences of viewing angle, subjects had to indicate their seat in relation to the presentation screen on a 3 (distance: close, midrange, far) × 3 (eccentricity: left, center, right) matrix.

The stimuli were presented in a resolution of 1024 by 768 pixels using an EPSON EMP 8300 video projector on a 5.6 m × 4.2 m silver screen. Each trial started with the visual and acoustic presentation of the question (e.g., "Which of both images would you prefer to have at home / do you like more?") followed by the image presentation. The presentation of the arrangement was counterbalanced; seven pairs contained the multi-perspective image on the left and for the remaining seven pairs it was shown on the right. Subjects' task was to evaluate which of the two images they preferred. Responses were collected by means of a questionnaire. Participants gave their response by marking the preferred answer with a cross. Stimuli were shown for 10 seconds and then replaced by a black blank for another 10 seconds in order to permit the participants to give their answers. The subsequent trial began with the presentation of the next question. The total duration for the study, including the introduction and the collection of the questionnaires at the end was approximately 20 minutes.

Results. Data from all questionnaires were analyzed using SPSS 14. Data was excluded from further processing if there was more than one answer, no answer or no clear answer could be coded. After this trimming procedure 91.5% of the data remained for further processing. The number of responses differed between the images within a range of 446 to 451 valid answers. Responses to the test images were further transcribed into two categories; multi-perspective ("multi") or single-perspective ("single") preferences per subject and image.

Figure 5 summarizes the responses for each image following this differentiation. For the control stimuli, where no differences existed between left and right images, responses to the left image were coded into "multi" and responses to the right image were coded into "single". This was an artificial classification, used only to enhance comparability and facilitate further processing. Frequencies of responses for each image were compared using chi-squared analyses. Although the number of multi-perspective choices varied between the images (n = 245–410) statistical testing revealed a reliable preference for the multi-perspective presentation for all images, $\chi^2(1, n \geq 446) \geq 3.94$, $p < .05$. Significance tests revealed no preference in the control images, $\chi^2(1, n \geq 446) \leq 2.58$, $p > .05$.

A further chi-square statistic was applied to examine the possible influence of viewing angle (seating position) on subjects' ratings. Participants were distributed relatively equally over the 9 possible positions, ranging from $n = 37$ (8.2%) to $n = 63$ (14%). For the majority of the images (11 test images, 1 control image) there was no effect of seating postion, $\chi^2(1, n \geq 445) \leq 14.55$, $p > .067$. However, for 3 test and 2 control images there was a significant effect, $\chi^2(1, n \geq 444) \geq 16.36$, $p < .05$. A closer look at this relationship revealed that sitting close and off center led to this effect.

Discussion. The common method of displaying three-dimensional scenes in computer graphics is based on perspective projection. Following the techniques of Renaissance painters in creating multi-perspective paintings, pairs of images with single and a multi-perspective versions were presented to the participants in Experiment 1. The results demonstrate a clear preference for images with multiple perspectives. These findings are in accordance with the limits of visual processing during a fixation. Since foveal (hence, sharp and colorful) vision is restricted to an area of approximately 2 degrees, the eyes have to move to explore the content of an image. In terms of computer graphics, this is analogous to a change in orientation of the virtual camera, while the position of the camera is stable. This reflects the way in which veduta painters constructed their paintings: keeping the same viewpoint, but changing orientation to the objects of interest.

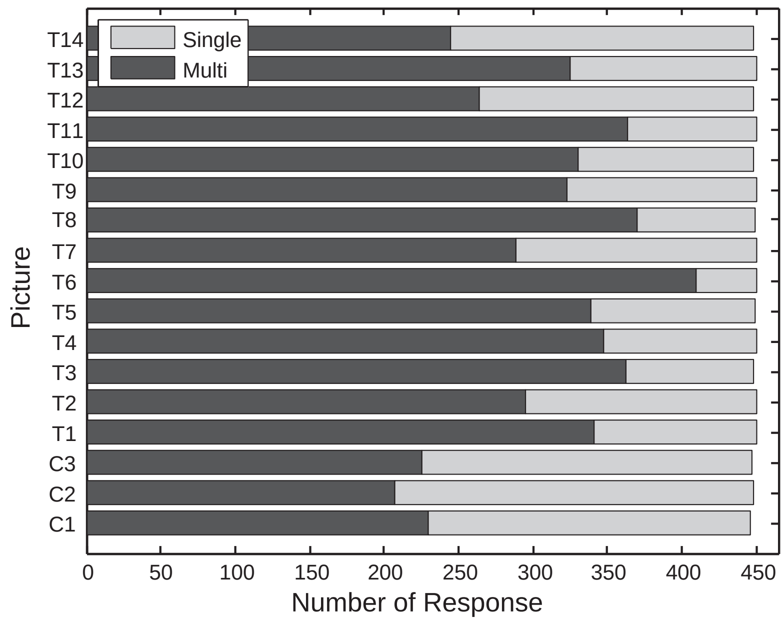

Fig. 5. Number of responses to the test images (T1-T14) and the control images (C1-C3).

The results of Experiment 1 would be weakened if the stimuli consisted of paintings only: A manipulated painting is unlikely to be as pleasing as the original, especially when perspective distortions were applied. However, we also found a preference for multi-perspective images in photographs and computer generated graphics. Hence, it can be seen that the preference was for the number of perspectives, regardless of the type of image. Analysis of angle of view (seating position) revealed a vulnerability of perspective perception. For the majority of the images, the position in the hall did not play a role in the aesthetic judgment. This is predicted by current visual psychophysics literature [Koenderink and van Doorn 2003]. However, the differences between original and manipulated images were obviously strong enough to elicit a reliable preference. For three of the test images and two of the control images, the position in the hall was a notable influence. In all of these five cases, the following criteria can be applied: Subjects seated in the front rows and with extreme viewing angle tended to have a bias towards the image presented on the same side, that is, sitting left from the screen biased towards the left image. Reanalysis of the relevant images did not indicate any other common features that could otherwise explain these results. Moreover, since this influence was also found for control stimuli, where both images were actually the same, it can be concluded that certain viewing angles can affect how the perspective projection is perceived.

Experiment 2 was designed to address the question of whether the differences in preference might also be reflected in eye movement patterns of observers.

3.2 Experiment 2

Participants. Twenty subjects (6 male, 14 female) participated in the eye tracking study, ranging in age from 20 to 41 years (mean 28.8 years, SD 6.8 years). All participants reported normal or corrected-to-normal vision.

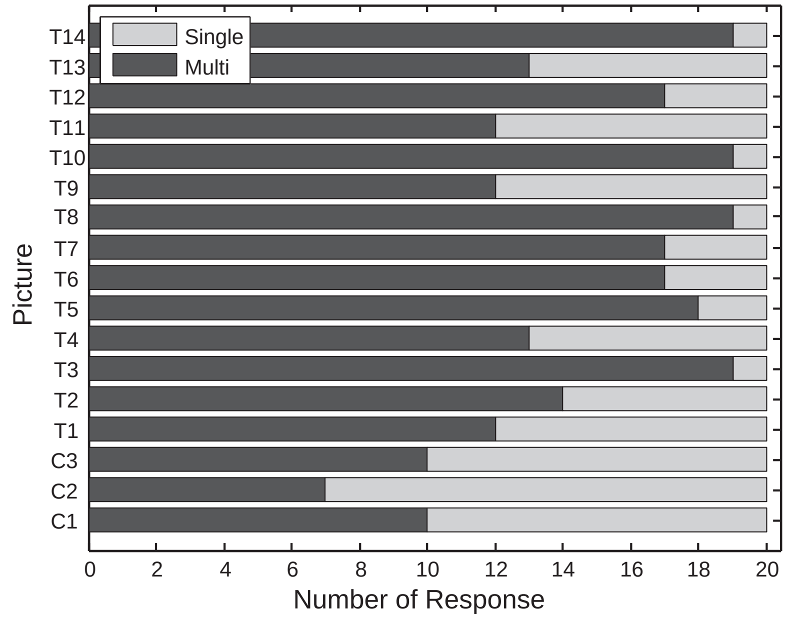

Fig. 6. Number of responses to the test images (T1-T14) and the control images (C1-C3).

Stimuli. The same images as in Experiment 1 were presented on a 19-inch LCD screen. With a screen to eye distance of 60 cm, the visual angle subtended 34.7 deg horizontally and 28.4 deg vertically. Eye movements were recorded at a sample rate of 120 Hz using the Eyefollower system, LC Technologies Inc.

Procedure. All participants were tested in a dimly lit laboratory in single sessions. Written consent was obtained and subjects provided answers to questions concerning demographic data. Subjects were instructed to use the "A" and the "L" key to indicate left vs. right responses, respectively. Following this, the eye tracking system was calibrated to the subject. All instructions were presented on the screen. After a training trial, subjects were given the opportunity to clarify any instructions. Seventeen trials were then presented. Each trial started with the visual presentation of the question (e.g., "Which of these images would you prefer to have at home / do you like more?") followed by the image onset. Stimuli were shown for 10 seconds and then replaced by a black blank screen for another 10 seconds in order to permit the participants to press a key. The subsequent trial began with the presentation of the next question. The total duration of an experimental session was about 15 minutes.

Results. First, we were interested in whether the results from Experiment 1 could be replicated in this smaller sample. Following the same categorization procedure used in the previous study, responses to the test and control images were differentiated into "multi" vs. "single". The responses per image are shown in Figure 6. Frequencies of responses for test and control images were compared using chi-squared analyses. For the test stimuli, we found a significant effect of number of perspectives on preferred image, $\chi^2(1, n = 280) = 93.73, p < .001$, whereas no reliable differences were found for the control images, $\chi^2(1, n = 60) = 0.60, p > .05$ (Figure 7(A)). Detailed chi-squared analyses on the preference for multi- versus single-perspectives in each test stimulus reached significance for 9 of the 14 test images. However, the tendency to prefer multiperspective images is apparent in all test images (Figure 6).

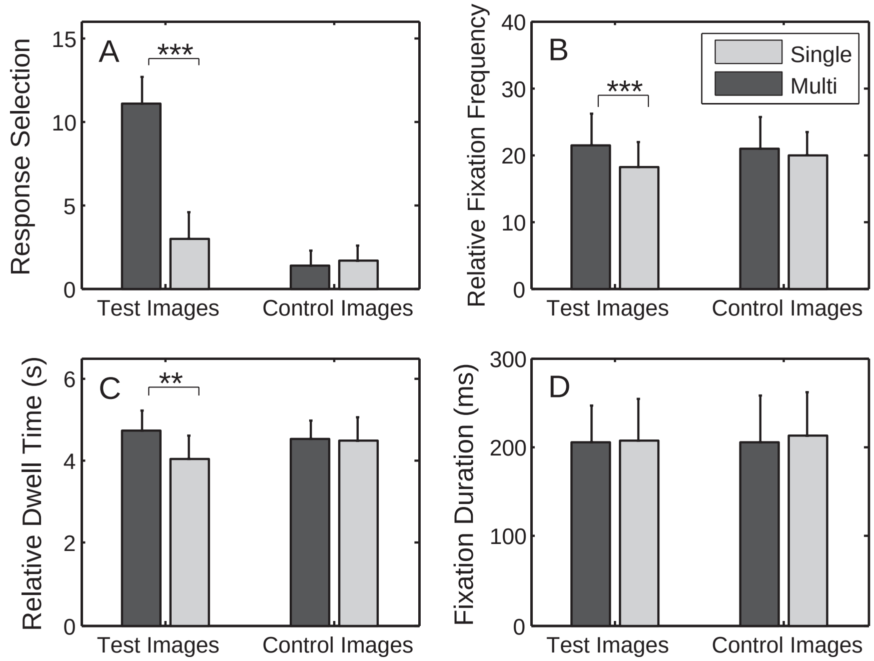

Fig. 7. Number of responses and gaze behavior. Selection of the multi- vs. single-perspective presentation for test and control images by key press (A), relative fixation frequency (B), relative dwell time (C) and the median fixation duration (D).

For the analysis of gaze behavior, we compared number of fixations, dwell time and fixation duration. The overall results of this analysis are shown in Figure 7. The frequency of fixations was found to be higher on multi-perspective images. This effect was statistically significant for the test stimuli, $\chi^2(1, n = 11072) = 71.54$, $p < .001$, but no difference was obtained for the control stimuli, $\chi^2(1, n = 2442) = 1.67$, $p > .05$ (Figure 7(B)). Single comparisons for each image revealed a significantly higher number of fixations on multi-perspective images for 8 of the 14 test stimuli, $\chi^2(1, n \geq 763) \geq 4.50$, $p < .034$, with no difference in number of fixations in each image in control stimuli, $\chi^2(1, n \geq 787) \leq 2.14$, $p > .144$. For the analyses of dwell times, fixation durations were summarized per image and subject and applied to a 14×2 repeated measures ANOVA, with stimulus and perspective (single vs. multi) as within factors. The reliable effect of the factor *stimulus*, $F(13, 247) = 2.25$, $p < .005$, $\eta^2 = .106$, indicates that dwell times were not equally distributed between the stimuli. The significance of the factor *perspective*, $F(1, 19) = 11.93$, $p < .005$, $\eta^2 = .386$, provides an explanation for this result: dwell times were longer for multi-perspective than for single-perspective images ($M = 4.72$ *vs.* 4.04 *s*; see Figure 7 (C). Moreover, we found a reliable interaction between both factors, $F(13, 247) = 1.82$, $p < .05$, $\eta^2 = .087$, demonstrating an interesting development of the distribution of dwell times over testing. At the beginning of the test session, dwell times were relatively equal for multi- and single-perspective images, but, for the latter stimuli longer, dwell times were clearly assigned to multi-perspective images. The control stimuli were similarly analyzed, applying the data into a 3×2 repeated measures ANOVA. No statistically relevant effects were found, neither for stimulus, $F(2, 38) = 1.70$, $p > .05$ nor for dwell time, $F < 1$ (see Figure 7(C)). No indication of interaction was found, $F(2, 38) = 1.16$, $p > .05$. Finally, fixation durations were analyzed with a 2×2 repeated measures ANOVA, with stimulus type (test vs. control) and perspective (multi vs. single) as within factors. No differences were found, neither for stimulus type, $F < 1$, nor for perspective, $F(1, 19) = 1.28$, $p = .272$; with no interaction, $F < 1$ (Figure 7(D)).

Discussion. Experiment 2 aimed to investigate the findings from the first study under controlled laboratory conditions and to elaborate on the findings through the registration of eye movements. First, the results from this experiment confirm the findings of our first study in demonstrating a general preference for multi-perspective images. This general effect was substantiated by analysis of number of fixations and dwell times. However, more thorough analysis of response and the allocation of fixations to each image type across each stimulus suggested that these differences were consistent across all test stimuli. One possible explanation for this is suggested by analysis of dwell times. Dwell times differed not only in relation to single- vs. multi-perspectives, but also between the stimuli. Moreover, the observed interaction supports the supposition that differences between the images evoked differential scanning and processing behaviors. However, no differences were found between the stimuli in relation to fixation durations, one of the parameters most likely to be sensitive to differences in presented stimuli. Thus, it can be concluded that despite possible questions regarding the heterogeneity of the stimuli, the multi-perspective presentation is nonetheless clearly preferred.

4. GENERAL DISCUSSION

In the present study, paintings, photographs and computer rendered images were shown in two different forms: images conformed either to the rules of perspective projection or the multi-perspective approach characterised by veduta painting.

In order to allow a direct comparison of these approaches, both forms of one image were presented simultaneously. In Experiment 1, the aesthetic judgments of the participants revealed a clear preference for the multi-perspective form. Control stimuli containing identical images emphasize this preference in that they reveal no preference. In Experiment 2, in addition to the aesthetic judgments, subjects' eye movement behavior was recorded as they explored the images. Again, judgments were biased towards multi-perspective images. Eye movement patterns revealed an increased number of fixations and longer dwell times on multi-perspective images. Fixation durations were analyzed since their relationship to depth of cognitive processing is well known [Velichkovsky 2002; Puolamaeki et al. 2005]. The analyses revealed similar fixation durations for the two forms of presentation and the control images, which leads to the conclusion that the general level of cognitive processing did not differ between the different perspective presentations.

Hence, it might be useful to understand the higher frequency of fixations and the longer dwell times on multi-perspective images in relation to the dynamics of visual attention. It is well understood that human attention is in itself a complex process, determinated by several levels of psychophysiological mechanism. At the lowest level, these mechanisms are allied with processing saliency, with saliency being spatially coded as a combination of local visual features of orientation, density, etc. [Itti et al. 1998]. Objects and areas of higher saliency in a scene attract more attention and are more likely to be fixated [Torralba et al. 2006]. Since the image pairs were quite similar in terms of their physical features, it can be assumed that saliency for each pair would be comparable. The discovered shift in number of fixations and dwell times towards the multi-perspective versions cannot be attributed the physical properties, but rather to higher level factors, such a meaningful composition of the scene, as in the clearly defined double semantic focus of Canaletto's "The new market in Dresden of Moritzstrasse" (see Figure 2). While the role of composition in active (attentive) perception and in the comprehension of visual scenary is not a new concept (cf. Arnheim [1998] and Velichkovsky and Hansen [1996]), this study demonstrates its relevance in aesthetic preferences in real and virtual environments. More indepth analyses is required to substantiate these findings and demarcate the practical implications for computer graphics.

This study further demonstrated, for the first time, a preference of multi-perspective images apparent not only in subjective judgments but also in gaze behaviors. These results have bearing on computer-generated environments, since the approach mentioned in 2.2 could be extended to fit presentation to subjects' preferences. The camera model and current human-based rendering methods could easily be adjusted to take account of these findings (see, e.g., Levoy and Whitaker [1990]). Although, in the present research, rendering methods were evaluated based on aesthetic judgments, we also expound the possibility of incorporating objective measures of ongoing user behavior in the design of computer-generated virtual spaces. With the current impetus for increasingly realistic virtual space, and the parallel focus on user behavior determined interface design, it would be particularly interesting to see how the approach of multiple perspectives would influence the perception of virtual space through eye movement analyses. Since the eye itself is analoguous to a rather low-grade camera obscura, the data from which is constantly rendered by the brain as a perceived perfect reality, there is much from visual perception that can be applied to the construction of subjectively gratifying virtual realities without the need for increased definition in the graphics. Enhanced immersion based on multi-perspective presentation could not only be achieved in computer games but also in other virtual reality applications. Stimuli compared in the current investigation consider the basic characteristics of human visual processing by presenting one vs. two perspectives. Eye movements under natural viewing conditions may lead to more numerous changes in perspective. However, it can be seen from this intial investigation that the introduction of another perspective in a static scene leads to an enhanced aestethic experience. Computer graphics and user-based interface design should benefit from applying what is known about about psychological processes of active vision and the methods that have been developed to measure and record processing in real time. The image of a reality, whether virtual or sensory, is the primary gateway to the experience of a real world. The information gathered by the eye and the information chosen as the eye moves must be the primary target if we are to achieve a dynamic adjustment of presentation based on realtime feedback from human behaviors [Velichkovsky and Hansen 1996].

Understanding images as interfaces to a meaningful world emphasizes the importance of design based on users' ongoing needs, interests and goal-directed activities. Eye tracking will be the method of choice for creating a dynamic relationship between synthetic stimulus and percept. We concieve of such an attention-focused approach as investigating the possiblities for realtime readjustment of projected perspective with respect to the user's gaze direction. We envisage a dynamic space where the user's point of perspective at every region or every object within the virtual space would change with every shift in visual attention and movement of the eyes, epitomizing the reality so effectively rendered by the human brain.

ACKNOWLEDGMENTS

The authors thank Fiona Mulvey and two anonymous reviewers for valuable comments to earlier versions of the manuscript. Thanks are due to Sascha Weber, Markus Joos and Michael Heubner for technical support in data acquisition and to Jan Schreier, Nia Katranouschkova and Claudia Geitner for the preparation of stimulus material.

REFERENCES

Agrawala, M., Zorin, D., and Munzner, T. 2000. Artistic multiprojection rendering. In *Proceedings of the Eurographics Workshop on Rendering Techniques 2000* (London, UK). Springer-Verlag, New York, 125–136.

Angel, E. 1997. *Interactive Computer Graphics—A top-down approach with OpenGL*. Addison-Wesley, Reading, MA.

Arnheim, R. 1998. *The Power of Center: A Study of Composition in the Visual Arts*. The Regents of the University of California, Berkeley, CA.

Dodge, R. 1900. Visual perception during eye movement. *Psych. Rev. 7*, 454–465.

EDGERTON, S. Y. 1975. *The Renaissance Rediscovery of Linear Perspective*. Basic Books, New York, NY.

FINDLAY, J. M. AND GILCHRIST, I. D. 2003. *Active Vision: The Psychology of Looking and Seeing*. Oxford University Press, Oxford, UK.

FOLEY, J. D. 1999. *Computer Graphics—Principles and Practice*. Addison-Wesley, Reading, MA.

FRANKE, I. S., RIEGER, R., AND ZAVESKY, M. 2005. The power of frustum. *Neue Medien in der Informationsgesellschaft 2*, 309–317.

FRANKE, I. S., ZAVESKY, M., AND DACHSELT, R. 2007. Learning from painting: Perspective-dependent geometry deformation for perceptual realism. In *IPT-EGVE Symposium*. IPT-EGVE, 117–120.

GROH, R. 2000. Romanik interaktiv. In *Vom Realismus der Bilder. Interdisziplinaere Forschungen zur Syntax bildlicher Darstellungsformen*, K. Sachs-Hombach and K. Rehkaemper, Eds. Scriptum Verlag, Magdeburg, Germany.

GROH, R. 2005. *Das Interaktionsbild—Theorie und Methodik der Interfacegestaltung*. TUDpress, Dresden, Germany.

GROH, R., FRANKE, I. S., AND ZAVESKY, M. 2006. With a painter's eye: An approach to an intelligent camera. In *The Virtual*. Stockholm, Sweden.

ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 20,* 11, 1254–1259.

JOKELA, T., IIVARI, N., MATERO, J., AND KARUKKA, M. 2003. The standard of user-centered design and the standard definition of usability: Analyzing iso 13407 against iso 9241-11. In *CLIHC '03: Proceedings of the Latin American Conference on Human-Computer Interaction*. ACM, New York, 53–60.

KLOTZ, H. 1990. *Filippo Brunelleschi: The Early Works and the Medieval Tradition*. Rizzoli Intl Pubns, Stuttgart, Germany.

KOENDERINK, J. J. AND VAN DOORN, A. J. 2003. Pictorial space. In *Looking into Pictures: An Interdisciplinary Approach to Pictorial Space*, R. S. H. Hecht and M. Atherton, Eds. MIT Press, Cambridge, MA.

LEVOY, M. AND WHITAKER, R. 1990. Gaze-directed volume rendering. In *SI3D '90: Proceedings of the 1990 Symposium on Interactive 3D Graphics*. ACM, New York, 217–223.

PALMER, S. E. 1999. *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA.

PUOLAMAEKI, K., SALOJAERVI, J., SAVIA, E., SIMOLA, J., AND KASKI, S. 2005. Combining eye movements and collaborative filtering for proactive information retrieval. In *SIGIR '05: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, 146–153.

RADEMACHER, P. AND BISHOP, G. 1998. Multiple-center-of-projection images. In *SIGGRAPH '98: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, New York, 199–206.

RAUBER, T. 1993. *Algorithmen in der Computergrafik*. B. G. Teubner, Stuttgart, Germany.

ROMAN, A., GARG, G., AND LEVOY, M. 2004. Interactive design of multi-perspective images for visualizing urban landscapes. In *VIS '04: Proceedings of the Conference on Visualization '04* (Washington, DC). IEEE Computer Society Press, Los Alamitos, CA, 537–544.

TORRALBA, A., OLIVA, A., CASTELHANO, M. S., AND HENDERSON, J. M. 2006. Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psych. Rev. 113,* 4, 766–786.

VELICHKOVSKY, B. M. 2002. Heterarchy of cognition: The depths and the highs of a framework for memory research. *Memory 10,* 5-6, 405–419.

VELICHKOVSKY, B. M. AND HANSEN, J. P. 1996. New technological windows into mind: There is more in eyes and brains for human-computer interaction. In *CHI '96: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, 496–503.

VON HELMHOLTZ, H. 1867. *Handbuch der physiologischen Optik*. Voss, Leipzig, Germany.

WATT, A. 1993. *3D Computer Graphics*. Addison-Wesley, Reading, MA.

YU, J. AND MCMILLAN, L. 2004. A framework for multiperspective rendering. In *Proceedings of the 15th Eurographics Symposium on Rendering (EGSR04)*. The Eurographics Association, 61–68.

ZOMET, A., FELDMAN, D., PELEG, S., AND WEINSHALL, D. 2003. Mosaicing new views: The crossed-slits projection. In *IEEE Trans. Patt. Anal. Mach. Intell. 25*, 6, 741–754.

ZORIN, D. AND BARR, A. H. 1995. Correction of geometric perceptual distortions in pictures. In *SIGGRAPH '95: Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*. ACM, New York, 257–264.

Received April 2007; revised October 2007; accepted October 2007