3D Display*

Introduction

Graphical human-computer interfaces applying windows, icons, menus and mouse pointers have considerably simplified the use of computer programs compared with the use of purely text-oriented input and output techniques. However, strictly 2D surfaces present obvious restrictions: when several applications run simultaneously, overlapping windows make it difficult to watch the screen and to follow its contents. In this context, 3D displays literally offer one more dimension for visualizing the data flow and the interplay of programs in complex multimedia applications in a very natural way.

Binocular vision allows humans to perceive and unambiguously interpret spatial structures without additional mental effort [1]. Thus, stepping into the third dimension could make it easier for users to perceive and understand complex information structures. In experiments on the understanding of abstract information networks presented in 2D versus 3D, Ware and Franck [2] found evidence that true 3D viewing can increase the size of a graph that can be understood by a factor of three. Compelling examples of the more effective use of the limited computer-screen space by 3D presentation were developed at Xerox PARC. For instance, Card et. al. [3] implemented an animated Cone Tree browser to three-dimensionally visualize the structure of hierarchical databases and Robertson et. al. [4] designed a Perspective Wall providing a detailed view of multimedia objects on the front part of the wall and context information on the perspective peripheral left and right wings (refer to [5] for an overview of approaches for coping with screen-space limitations).

To fully exploit the potential of 3D visualization, a further step forward in the evolution of human-computer interaction has been suggested [6]: While current computer operating systems like MS Windows, MacOS and Unix implement command-based, direct-manipulation interfaces [7], next-generation user interfaces should supplement this concept by introducing non-conventional controls [8] and intelligent interface agents to support non-command interactions [9]. The idea is to give the computer sensors so that it can constantly observe the user. Being aware of the user's activities and taking the situational context into account, the interface agent will be able to interpret the user's intentions and to infer how to optimally adapt interaction to the user's needs. Hence, the interface agent could relieve the user of "routine" actions, giving him or her the freedom to fully concentrate on the task at hand – rather than on operating the computer.

In previous studies gaze tracking has proven to be a powerful approach when implementing non-command interactions [10]. Knowing the user's current point of fixation on the display screen and the immediate history of eye movements allows the interface agent to distinguish whether the user is focusing attention on a particular item (e.g. an icon representing an application program) or whether he or she is unintentionally scanning the display screen [9] . The fundamental studies on vision-based non-command interactions by Jacob [11] demonstrated that when the system responds quickly and accurately users will forget about the fact that the computer is reading their eyes – they get the feeling that the system anticipates their intentions before they express them.

The approach outlined so far can be summarized as an attempt to merge the benefits of 3D visualization techniques with computer vision in order to make interaction with a computer more user-friendly. Up to now, experiments concerning the measurable advantages of this approach have been very sparse and it is impossible to give reliable conclusions about the user's appreciation of the new concept and the usability of the interface as a whole.

* Source: http://atwww.hhi.de/~blick/Introduction/introduction.html

Against this background the authors of this paper have developed a practical testbed in order to perform usability tests and to asses cost/benefit tradeoffs in an anticipated application scenario. A prototype system [12] was used for preliminary user testing and shown to the public at an international broadcasting exhibition. At present, the testbed's key components are free-viewing 3D displays, a visual operating system providing a graphical 3D user interface, a camera to sense the user's head position and motion (head tracker), and equipment to measure the user's point of fixation (gaze tracker). The 3D displays eliminate the need for 3D viewing glasses and provide the users with high-resolution images, including live stereo-images for video-conferencing. With the head tracker, a simple movement of the head is sufficient to open the view of a document hidden behind a visually overlapping foreground object (dynamic perspective). Simultaneously, the gaze tracker determines the current point of fixation, so that looking at the formerly hidden document will pull it closer to the user making it easier to read. Moreover, the gaze tracker adjusts the process of image rendering so that only the object being looked at appears in full focus – objects out of the user's gaze are temporarily considered unimportant and therefore shown out of focus, helping them to fade from perception (active accentuation). This effect mimics the limited depth of focus of the human eye [13], [14]. Commercial voice control software as well as a video-based hand tracker for direct manipulation of visualized 3D objects [15] will be integrated in the testbed in order to offer a variety of multimodal interactions.

3 D Visual OS

The core element of the proposed user interface is a new operating system based on the concepts of both object-oriented programming (in terms of adaptation and the use of already available program functions) and visual programming (in terms of the display of software modules). This means that users can use a graphic editor to create and "program" their environment and applications by simply linking appropriate software modules together. Compared to conventional visual programming tools, the proposed system builds on the advantages of 3D visualization, allowing a clearly structured representation of the interconnected program modules. The operating system runs on a Silicon Graphics Onyx computer and uses the virtual reality software dVS by Division Ltd. as a basis for generating 3D graphics and for implementing user interactions.

The operating system subdivides all user-accessible software modules into three levels of complexity, comprising *primitives* as the smallest units (basic arithmetic or graphic functions), *components*, and large-scale *applications* (aggregations of lower-level modules), respectively. The graphic representations of primitives (called gadgets) can be used to add animation, sound or database queries to a module, for example. The gadgets are stored in the library of the VOS and can easily be modified in order to change their visual appearance, without changing their particular functionality. Fig. 2 shows representations of a gadget used for scrolling the visible part of a text document. Several gadgets may be combined in order to form a component which can be used in various applications. In order to form larger software entities from lower-level modules, a module may have "docks" which can be visually connected via a pipeline in order to enable the exchange of information packages between the modules (Fig. 3). Thus a network of interconnected software modules can be created, tested and modified step by step in order to ultimately build a complex application program.
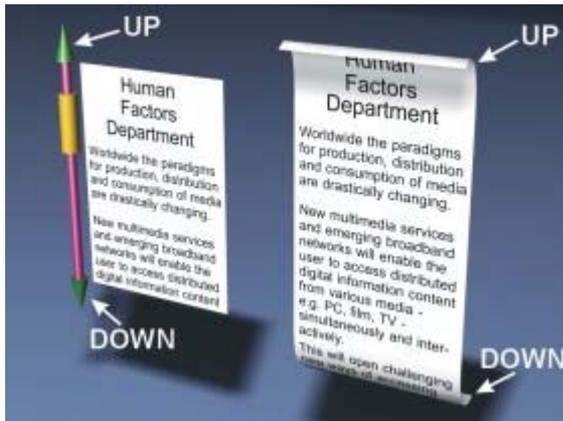
Fig. 2: *The appearance of a basic gadget is easily adapted to the user's preference without changing its functionality (scrolling the contents of a text document with a conventional scroll-bar vs. applying a papyrus-scroll metaphor).*
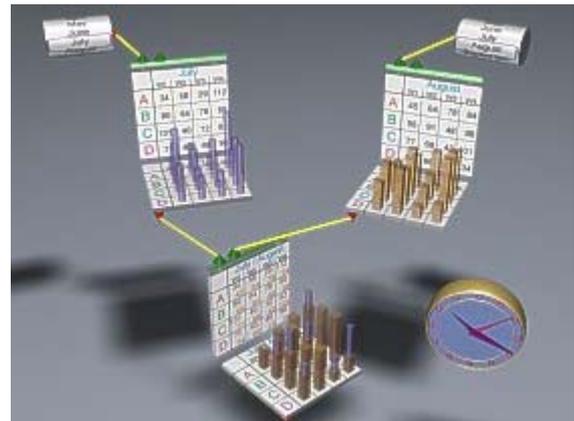
Fig. 3: *The user-accessible software modules provide synapses (docks). Pipelines between the docks enable the transfer of information and allow the user to create complex applications (in this example a multi-dimensional spreadsheet) by visually connecting the docks of basic components.*

In order to get a clearly structured representation of complex programs, users can apply a "zoom-out" function. This way, components can be compiled to form an aggregation where only the external docks are visible and the inner network is hidden (Fig. 4). It is also possible to zoom-in in order to visualize the large number of primitives and components forming the network of software modules in an application program [6].
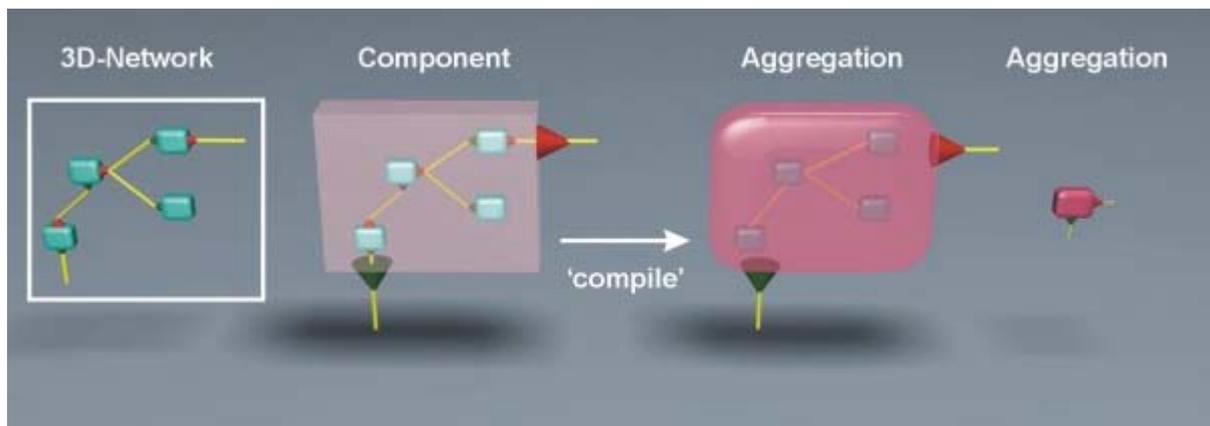

Fig. 4: *The network of software modules forming an application can be visualized at different levels of detail using a particular zoom function. At each level, the accessible external docks are accentuated. Obscuring the inner network improves the clearness of the presentation.*

When talking of "users" in the framework of the new operating system traditional differences between software developers and end users are intentionally diminishing. Future end users should have the possibility to set up their own individually tailored application programs. Instead of using a huge monolithic program with a multitude of possibly never used functions, end users should be allowed to combine software modules from different suppliers according to their particular needs. The VOS aims

at making interchangeable software components applicable to end users whilst providing an easy-to-use, clearly structured visual interface in connection with intuitive manipulation techniques.

Application Scenario

The boxes in Fig. 5 show icons of frequently used application programs or tools mapped onto their surfaces. By changing the viewing position (head movement), the user will see the boxes from different perspectives. The boxes change their perspectives overproportionally, so that it is possible to make each side visible by small head movements. Now imagine that the user is looking at one of the application icons (e.g. a spreadsheet application). Since the computer knows the gaze point, an "event" is triggered, starting an animation primitive which magnifies the fixated spreadsheet icon for feedback and improved visibility. If the user continues to look at the icon for a certain period of time (e.g. for 150 ms, as proposed in [11]), thus signaling an increased interest, the interface agent will instance the spreadsheet program. A sequence of visual interactions will allow the user to select and visualize numerical data in a 3D bar diagram. In the example shown in Fig. 3, some data related to the months of July and August are displayed on a weekly basis. Looking at the corresponding input and output docks of the spreadsheets will create pipelines and produce a combined presentation of the data. Meanwhile, the operating system will automatically re-position the applications within the limited 3D display volume to warrant optimal visibility. The position and size of an object can also be manually relocated by drag-and-drop operations with a conventional 2D input device, such as a mouse. In this case, the user's gaze selects the object to be manipulated in a predefined mode (movements in a plane parallel to the screen or along the z-axis, respectively).
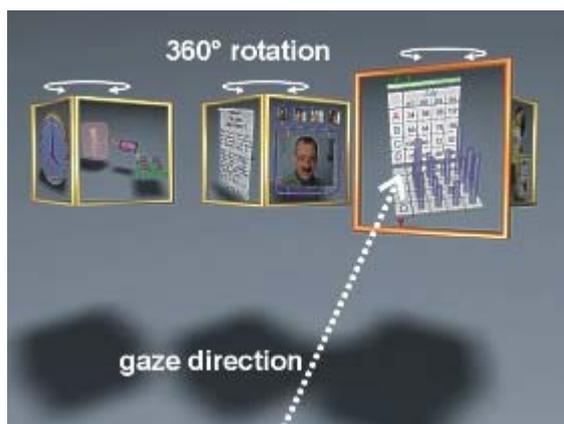

Fig. 5: *Tool boxes showing icons of applications or documents change perspective in response to the user's head movements and react when the user looks at a particular icon by, e.g., launching the relevant application program.*

Another scenario requires the user to find the "music box" icon located on the surface of one of the tool boxes and to launch it by eye-controlled interaction. After selection, the interface agent will automatically position the application in the displayed 3D volume. When the user looks at the music box, it comes closer and opens its cover. Next, the user can visually select a title and start playing it by glancing at the virtual play button. After playing and when no further interactions happen, the interface agent will remove the music box from the display.

Fig. 6 and Fig. 7 show the implementation of a local file browser and a Web browser, respectively. In Fig. 6 the object in the background represents a file system which is connected to a file-type filter (the

cylindrical object with buttons to select the type of file desired). When the user looks at a particular file, the corresponding viewer (in this example a text viewer) will be launched. The text automatically scrolls when the user's gaze point has reached the top or bottom of the text page. Glancing at a hypertext item on a Web page (Fig. 7) for a certain dwell time automatically downloads the hyperlinked document. The previously loaded documents will move backwards, thus indicating the search path to the current document. Looking at a background document will in turn move it closer. Changing the viewing position by moving the head discloses occluded documents and thus helps the user keep track when browsing the Web.
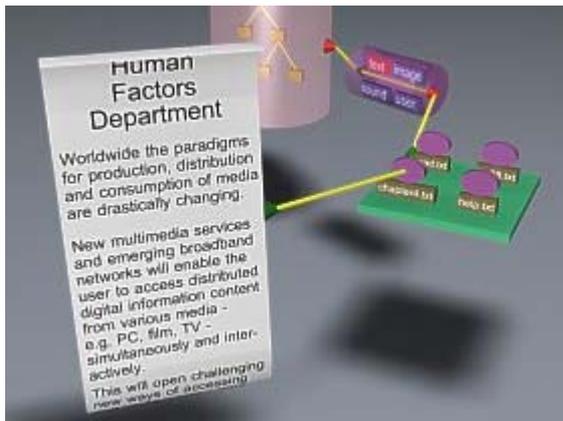

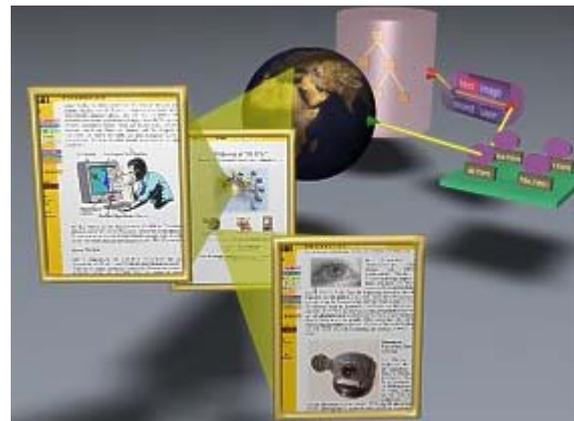Fig. 6: *Implementation of a local file browser.*

Fig. 7: *Implementation of a Web browser.*

3 D Display

Because the computer evaluates the user's gaze, the proposed system requires a 3D display without polarizing glasses or any other head gear occluding the user's eyes. Autostereoscopic 3D displays [16] are based on the concept of directional multiplexing, which means that the different perspective views are visible only from a limited number of fixed viewing positions. When the user positions his or her head so that the eyes are within the viewing zone, both views are immediately fused to create the illusion of a 3D space. For practical reasons such displays must have a tracking mechanism in order to optically address the eyes, both at a fixed head position and also when the user moves.

Fig. 8 shows the optical principle of a 3D display which we made in cooperation with Carl Zeiss (Germany). This display uses a movable lenticular screen in order to optically address the eyes over an extended viewing zone. The lenticular screen (made by Philips Optics, The Netherlands) is placed in front of an LCD screen. The left and right image contents are simultaneously displayed in columns side by side. As a result, columns 1, 3, 5, 7, etc. (labeled "R" in Fig. 8) display the information for the right eye, while columns 2, 4, 6, 8, etc. (labeled "L") display the information for the left eye. Since the lenticular screen has a directional selectivity in the horizontal plane, the color primitives of the LCD panel have to be aligned vertically one above the other in order to avoid color separation of the RGB components. Since the color primitives in commercial LCD panels are aligned horizontally, the LCD panel is rotated by 90 degrees. The lenticular plate separates the two stereo pictures for the viewer's eyes. Depending on head movements, the lens plate is mechanically adjusted to the left and right as well as in the frontal direction. The maximal tracking range is mechanically limited to about 30 cm for lateral and frontal head movements with a nominal viewing distance of 67 cm (this translates to a lens shift of 0.042 mm per 1 cm of frontal or horizontal head movement at the nominal distance; for more details refer to [17]). As the screen is composed of vertically oriented cylindrical lenses, vertical head movements do not require lens shifting; they are only limited by the tracking range of the head position sensor.
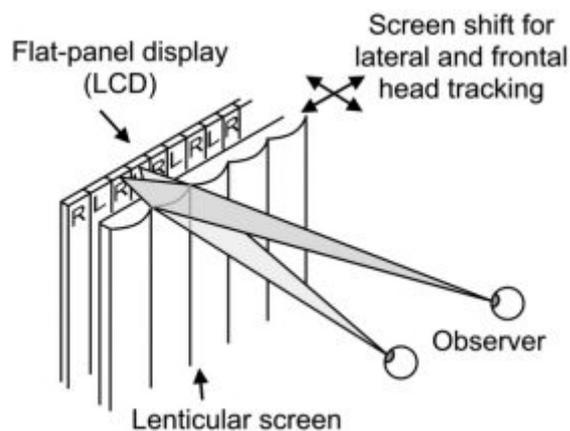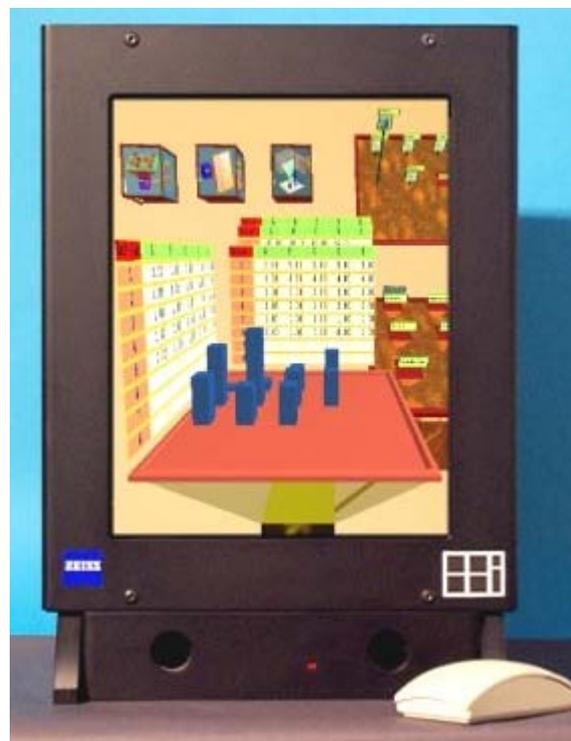

Fig. 8. *A lenticular screen allows the separation of viewed 3D images created in a raster-scan mode. The photo shows a 14-inch XGA resolution prototype 3D display developed for desktop applications.*

A 50-inch, projection-type 3D display has also been developed in cooperation with Philips Optics and

Cybertron (Germany). This display uses a dual lenticular screen with 1000 lenses and two LCD projectors, each with XGA resolution (Fig. 9). The cylindrical lenses of the back lenticular screen are used to form an array of left-right image stripes on an intermediate diffusing plate. The front lenticular screen has the same lens pitch as the back screen which causes the two stereo half images to be channeled to the left and right eyes. Again, the front lenticular screen is shifted mechanically in the lateral and frontal directions in order to follow the head movement (0.064 mm per 1 cm of head movement). The tracking range and the nominal viewing distance are about twice as large as for the desktop display. Both displays are single-user displays and provide changes in perspective (generated by the computer) in response to horizontal, frontal and vertical head movements within the tracking range.
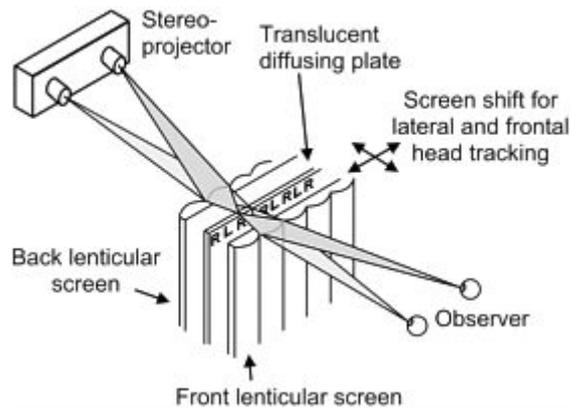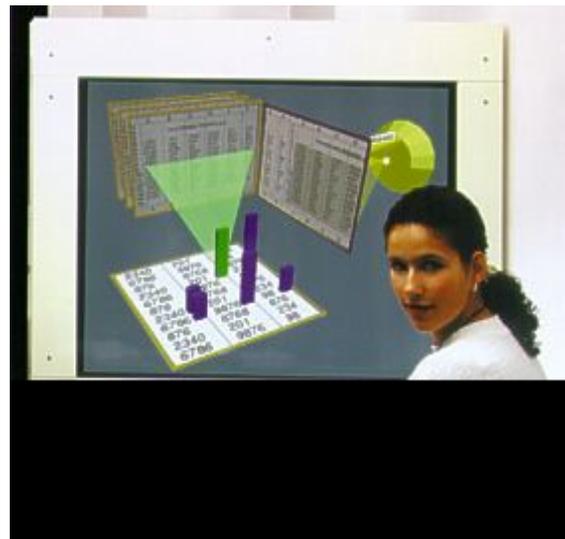

Fig. 9. *Principle and prototype of a 50-inch projection-type dual lenticular screen 3D display.*

The 3D displays developed so far have an inherent shortcoming which strains the eyes: the user must focus on a fixed viewing distance (the screen distance) although the stereo objects may appear close to the eyes or far behind the screen. By contrast, in natural viewing conditions the accommodation distance changes in accordance with the distance of the object observed (convergence distance of the lines-of-sight). To make display viewing more comfortable, we propose the "depth-of-interest" display [18] shown in Fig. 10. Note that the stereo images are not projected onto a physical display screen at a static location; instead the display provides a movable image plane enabling fixation-point dependent accommodation. The stereo images appear as aerial images floating in front of or behind a large Fresnel-type field lens. The location of the image plane is controlled by motorized adjustments of the projection optics (focus) in such a way that the aerial image appears at a distance corresponding to the stereoscopic distance of the object the user is looking at. This type of display needs to "know" both the user's head position as well as the stereoscopic depth of the currently fixated object (as estimated from the head position and gaze direction). As the viewer accommodates on the aerial image plane, accommodation distance and convergence distance coincide like in natural vision. Additionally, the display concept encompasses a natural depth-of-focus effect by depth-selective, spatial low-pass filtering of the projected images. As opposed to the lenticular screen displays, the "depth-of-interest" display has not yet been built; however, a proof-of-concept test using static stereo images was successful.
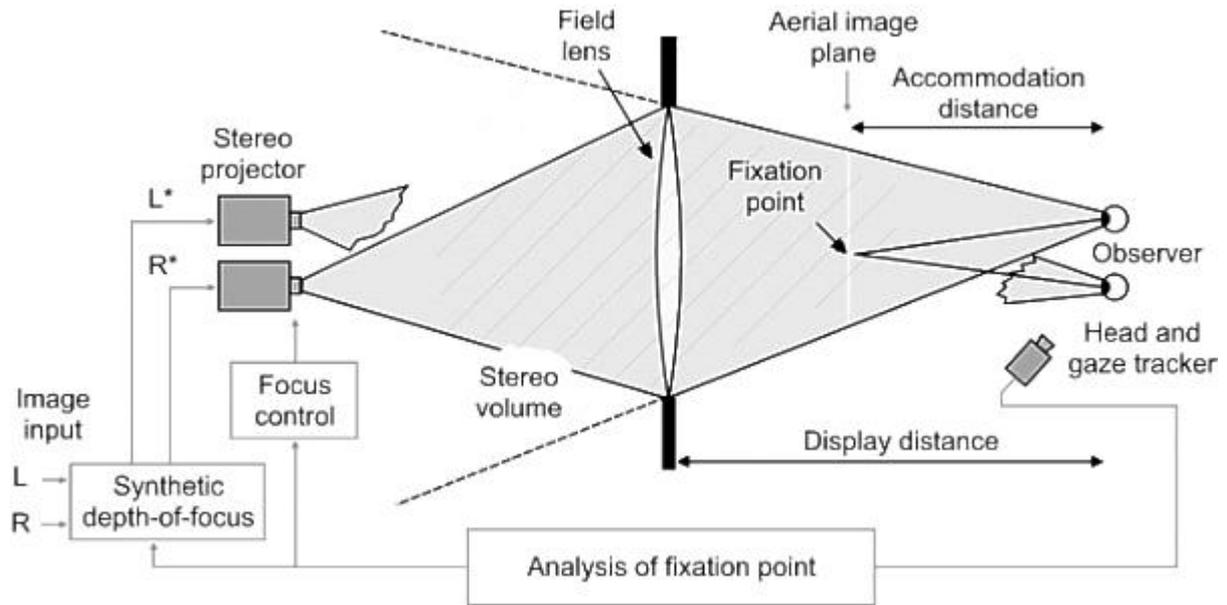
Fig. 10: *Components of the proposed depth-of-interest display: The position of the aerial image of the observed object corresponds to its apparent location within the stereo volume.*

Stereoscopic Image Representation with Synthetic Depth of Field

W. Blohm, I. P. Beldie, K. Schenke, K. Fazel, S. Pastoor
Heinrich–Hertz–Institut für Nachrichtentechnik Berlin GmbH
Tel: +49 30 31002-387 / Fax: +49 30 31002-213 / e–mail: blohm@hhi.de

Abstract - A novel concept of stereoscopic imaging providing a depth–of–field blur effect close to that experienced in natural vision is presented. With this concept, only a mini–volume of the reproduced three–dimensional space is displayed in full spatial resolution. This volume is centered around the current point of fixation of the viewer. In an exploratory study, the proposed concept of a so–called "depth–of–interest" (DOI) display was evaluated. Subjects assessed computer–generated 3D images with different depth–selective filtering characteristics (i.e., varying in–depth extents of the mini–volume with full resolution). Results indicate that the intended DOI display can indeed improve 3D viewing comfort. Apparently, a somewhat smaller depth of field than experienced in natural vision is preferred in stereoscopic image representations.

Keywords - depth–of–field simulation, stereoscopic display, 3DTV, depth–of–interest display, viewing comfort, human factors

Introduction

Due to the limited depth of focus of the human eye, only a certain depth range in front of and behind the current point of fixation is viewed in full sharpness. The center position of this depth of field (DOF) is linked to the fixation–point distance in natural vision (known as accommodation–convergence correspondence). As a result, the disparate retinal images of contours far behind or in front of the fixation plane are blurred, and in turn, no strong stimuli arise to fuse them. The visual system is thus able to cope even with excessive disparities without developing eyestrain.

Current stereoscopic display techniques require the viewer to decorrelate accommodation and convergence mechanisms [1]. Irrespective of the current point of fixation, the eyes have to accommodate on the display screen in order to perceive sharp images. Provided the DOF of the displayed stereo images covers the scene's full depth range, all depth layers are imaged in full sharpness on the retinas of an observer's eyes. This way, excessive disparities are perceived without any blur, giving rise to annoying double images. Hence, viewing 3D pictures with stereoscopic displays is likely to create visual strain when the disparity magnitude exceeds a certain limit[2,3]. Several proposals to avoid this strain have been reported in the literature, e.g., limited parallax ranges[4] or large viewer–to–screen distances [5]. However, there are a lot of 3D viewing situations feasible (e.g., desktop multimedia applications) where it is not possible to impose the proposed restrictions.

In order to overcome the outlined limitation of stereoscopic displays and allow for comfortable, non–straining spatial viewing, helpful indications were provided by a recent study of Wöpking[6]. His results suggest that viewing comfort might be improved by simulating the DOF effect of natural vision in stereoscopic images. We propose the term "depth–of–interest" (DOI) display for displays mimicking a DOF effect.

Based on a thin lens model of the optics of the human eye, a post–filtering technique was developed for simulating a DOF effect in video images. Using pixel depth (e.g., z–buffer data of computer–generated images) and fixation–point distance as input data, the method provides images appearing to have a DOF by depth–selective spatial lowpass filtering of images which are in or nearly in focus over the total depth range imaged.

In order to find an optimum DOF for stereoscopic image representations at DOI displays, several DOF–filtered versions of computer–generated experimental test imagery were rated by subjects with respect to the viewing comfort experienced. The in–depth extent of full–sharpness volumes around the point of fixation and the blurriness of depth layers further behind or in front of this point varied in

the presented test image versions. This was achieved by changing the aperture diameter for which the synthetic DOF was calculated.

It was expected that the optimum DOF would be the one produced with an aperture diameter comparable to that taken by an average eye pupil under the illumination condition given in the experiment. Smaller diameters might produce too little blurring of contours far behind or in front of the fixation plane giving rise to eyestrain, while larger diameters were assumed to cause visual discomfort because of too much image blur due to the smaller than usual depth of field.

Post–Filtering for DOF Simulation

When neglecting occluding boundaries, the formation of an image by a lens system (e.g., the optics of the human eye) can be considered as a linear optical process. Thus, a DOF effect may be simulated by linear filtering of images with indefinite or nearly indefinite DOF as proposed by Potmesil and Chakravarty[7].

Assuming a geometric optical model (i.e., aberrations and diffraction ignored), Figure 1 illustrates how rays originating from an out–of–focus object point "O" are focussed on a point "o" in image space and then spread out onto the retina forming a so–called circle of confusion rather than a single point[8]. The radius e of this circle increases with the amount of focus error, decreasing the apparent sharpness of an image.

It follows from Figure 1 (similar triangles in image space) that:

$$\varepsilon = (P/2) \cdot \left(\frac{b}{a} - 1 \right)$$
. (1)

Taking into account the thin lens formula (i.e., $1/f = 1/z + 1/a$ and $1/f = 1/\bar{z} + 1/b$ where f is the focal length of the lens), we get:

$$\varepsilon = b \cdot (P/2) \cdot \left(\frac{1}{\bar{z}} - \frac{1}{z} \right)$$
. (2)

To simulate a DOF effect when viewing images of a scene, points in the displayed image have to be spatially expanded as a function of distance z and the viewer's fixation-point distance $\bar{z}$ in accordance with Equation (2). A measure e ' of this expansion is derived by projecting e onto the display plane as outlined in Figure 2, where perfect imaging is assumed since the viewer's eyes accommodate to the display distance d.

From Figure 2 we obtain that $\varepsilon' = \varepsilon \, (d/b)$, so that we may write with Equation (2):

$$\varepsilon' = (P/2) \cdot d \cdot \left(\frac{1}{\bar{z}} - \frac{1}{z} \right)$$
(3)

The circles defined by e ' may be considered as point spread functions

$$h(x,y,\varepsilon') = \frac{1}{\pi \cdot \varepsilon^2} \begin{cases} 1 & \text{if } \sqrt{x^2 + y^2} \le \varepsilon' \\ 0 & \text{otherwise} \end{cases}$$
(4)

of the optical system to be simulated for each pixel position (x,y) in the image (see Figure 3). Thus, intensity images $I_{DOF}(x,y)$ which appear to have a DOF are obtained by linear filtering

$$I_{DOF}(x,y) = I(x,y) * h(x,y,\varepsilon') \qquad (5)$$

where $I(x,y)$ is supposed to be an image which is in focus or nearly in focus over the whole depth range imaged, and $*$ denotes convolution.

A deficiency of the presented post–filtering approach is that it does not correctly simulate focus effects near occluding edges. Thus, contours of foreground objects may lack appropriate blurring. To overcome this shortcoming, we applied the filter mask h obtained at the surface of an occluding object from Equation (4) just a few pixels beyond the object boundary as well, provided the occluding object is less in focus than the surrounding surfaces.

Subjective Evaluation of the Synthetic DOF Effect

Method

Subjects

The sample comprised eight subjects (five male and three female, between 25 and 35 years of age) who had normal or corrected–to–normal vision (visual acuity: Landolt test ³ 1.0) and were able to perceive depth from binocular vision (stereo acuity: Zeiss Polatest 100%). All subjects were familiar with stereoscopic image representation, but were naive as to the research question under investigation.

Technical Set–up

A mirror stereoscope, as shown in Figure 4, was used to display the experimental test imagery. Right and left–eye images were displayed at a viewing distance of 70 cm on two identical 21" (53 cm) monitors (Nokia 445X) with a resolution of 1280 × 1024 pixels and a refresh rate of 80 interlaced fields per second. The mean luminance produced by the monochrome test pictures was about 45 cd/m^2.

Stimuli

The stimuli were DOF–filtered versions of two static images and a full–motion image sequence. Both, the original static images and the original motion sequence were acquired from the same 3D test scene. As seen from Figures 5 and 6, this computer–simulated scene is composed of a quadrilateral platform located in front of a textured background, and several more or less futuristic objects positioned in different depth at or above the quadrilateral platform. An ellipsoidal object, serving as a fixation target, is flying along a curvilinear path back and forth through these objects. Its velocity varies smoothly along the trajectory. Horizontal movement is maximum (approx. 9 deg/sec) near Object 5 and approaches zero towards both turning–points.

From this scene, the original static images as well as the 204 frames of the original motion sequence were rendered for each eye point with infinite DOF using standard computer graphics software. The test scene was reproduced stereo–geometrically in unit scale so that the front face of the quadrilateral platform appears to be about 17 cm in front of the display plane, while the apparent position of the character pattern at the scene's background wall is about 32 cm behind the display plane. Considering the viewing distance of 70 cm and an interpupillary distance of 64 mm, this gave parallaxes running from –100 to +100 minutes of arc.

DOF–filtered imagery was created by image processing of the original static images and the frames of the original motion sequence, respectively. Using the post–filtering method outlined in the previous section, imaging by a lens with aperture diameters of P = 2.5 mm, 3.75 mm, 5.0 mm, 7.5 mm, and 10.0 mm was simulated. Since the flying object served as a permanent fixation target throughout the experiment, its z– position was assumed as fixation–point distance $\bar{z}$ in all DOF–filtered images. Including the non–filtered original imagery there was a total of six DOF conditions.

Figure 7 illustrates the effect of synthetic depth of field by showing DOF–filtered right–eye image versions of both static images employed in the experiment. In both, the flying object is at extreme positions. In the static image called "front position", Figure 7(a), it is at its very front position (about 15 cm in front of the display plane) partly occluding one of the backmost objects of the 3D scene (Object 8 in Figure 6), whereas the flying object is at its very back position (about 29 cm behind the display plane) in the static image called "back position". As seen from Figure 7(b), it is then partly occluded by one of the frontmost objects (Object 2 in Figure 6).

Procedure

The experiment consisted of two parts. In the first part, the subjects had to assess six DOF conditions of both static images. The full–motion test sequences, with the same six levels of DOF, were rated in the second part. A complete session lasted about 180 min. This included a break of at least 15 min between both parts of the experiment.

Static images. After presentation of 20 image pairs for training purposes (among others, the subjects were trained to concentrate on the flying object as a permanent fixation target), DOF–filtered versions of the same static image were presented as 6 × (6 – 1) = 30 pairs in randomized order to each subject (paired comparison method). Pairs of static image "front position" alternated with pairs of static image "back position". Both images of a pair were displayed sequentially, each for 10 sec with an intermediate gray image presented for 4 sec. The subjects were asked to give their rating of experienced difference in viewing comfort (second presented image of a pair relative to the first one) on a continuous comparison scale (ranging from "much less comfortable" to "much more comfortable" with intermediate points labeled "less comfortable", "slightly less comfortable", "the same", "slightly more comfortable", and "more comfortable"). The mean duration of this part of the experiment, including three breaks of up to 15 min each, was about 135 min.

Motion sequence. In the second part of the experiment, the subjects assessed the full–motion sequence with six levels of DOF. After a training phase with all six sequences, the test conditions were presented to each subject in a sequentially balanced Latin square order (presentation time of one sequence: 20.4 sec = 4 × 204 frames as loop / 40 full frames per sec). All sequences were presented a second time to each subject using a different sequentially balanced Latin square order (factor "replication"). After a sequence was displayed, the subjects gave their rating of absolute impairment of viewing comfort experienced during the presentation of the motion sequence. A continuous version of the five–grade CCIR/ITU–R impairment scale[9] (ranging from "imperceptible" to "very annoying" with intermediate points labeled "perceptible, but not annoying", "slightly annoying", and "annoying") was used. This part of the experiment lasted about 25 min.

Results

Static Images

The comparison ratings given by the subjects were mapped into numbers ranging from -3 ("much less comfortable") to +3 ("much more comfortable"). They were evaluated separately for each static image using an ANOVA for paired comparisons ratings (see Ref. 10, pp. 145–148). In these evaluations, equally spaced intervals for scale values and factor levels were assumed. The main effect is

significant for "front position" (F(5,210) = 6.10, $p <$ 0.05) as well as for "back position" (F(5,210) = 11.12, $p < 0.05$).

The results are illustrated in Figure 8. For both images, viewing was more comfortable with larger aperture diameters P, reaching a saturation level near $P = 5.0$ mm. The least significant differences, indicated by "yardsticks" in Figure 8, were calculated post–hoc using Tukey–HSD test (significance level 0.05). For the "front position" static images, differences in viewing comfort between condition $P = 0.6$ mm and all conditions above $P = 2.5$ mm are significant while ratings given for aperture diameters above $P = 3.75$ mm differ significantly from those of conditions $P = 0.6$ mm and $P = 2.5$ mm for the "back position" static images.

Motion Sequence

Two factors, "aperture diameter" with six levels and "replication" with two levels, were considered in the analysis of the motion sequence part of the experiment. Ratings were mapped into numbers ranging from 1 ("very annoying") to 5 ("imperceptible"). Assuming equally spaced intervals for scale values and factor levels, an ANOVA for repeated measurement designs was carried out. Only the main effects are significant ("aperture diameter": F(5,35) = 3.26, p < 0.05; "replication": F(1,7) = 12.69, $p < 0.05$).

With respect to the factor "replication" the maximum difference in mean values of the same aperture condition between first and second presentation was less than 0.6 units on the impairment scale. On average viewing comfort was rated 0.2 units worse for the second presentation. Means and standard errors calculated over subjects *and* both presentations are plotted in Figure 9. The plot exhibits a decrease in viewing comfort with increasing aperture diameter P beginning near $P = 5.0$ mm. Post–hoc analysis (Tukey–HSD, significance level 0.05) indicated that means for test conditions $P = 0.6$ mm, 2.5 mm, 3.75 mm, and 5.0 mm differ significantly from that of condition $P = 10.0$ mm.

Discussion and Conclusions

A DOF effect close to that experienced in natural vision was simulated in computer–generated stereo image pairs. The independent variable for this simulation was the lens aperture diameter for which the synthetic DOF was generated. This parameter determines the in–depth extent of the volume of full sharpness around the plane of fixation, and the degree to which depth layers more distant from this plane become blurred. Because of a lack of reliable devices for measuring the lines of sight of the viewer's eyes, a flying object served as a fixation target throughout the experiment. The subjects were asked to concentrate on this target.

Concerning the ratings of the DOF–filtered full–motion stereo sequences, the study confirmed the expected decrease in viewing comfort above a certain aperture diameter ($P >$ 5.0 mm in Figure 9). Though there exists a small tendency for an optimum near an aperture diameter of $P = 3.75$ mm, we encountered no drastic increase in impairment towards smaller apertures. This might be attributed to motion blur (contours of fixed objects appear unsharp[11] and, thus, annoying double images are suppressed), and to the fact that time intervals to induce any visual discomfort in positions with excessive disparities are quite short.

As outlined in Figure 8, however, results are the other way round when positions of the fixation target associated with excessive disparities are viewed as static images: Viewing comfort is less for smaller apertures, becomes highest (Figure 8(a)) or nearly highest (Figure 8(b)) for aperture diameters around P = 5.0 mm, but shows no decrease at larger aperture diameters. A possible explanation for the absence of this drop might be that viewers are accustomed to viewing photographs with rather small fixed DOFs. Differing results towards higher aperture diameters are expected, however, with a true DOI display (i.e., viewers are allowed to randomly fixate points in stereoscopic space and the DOF follows appropriately).

The results of both parts of the experiment suggest that the intended DOI display will allow a more comfortable and non–straining viewing of stereoscopic images. A substantial extension of the exploitable depth range at stereoscopic displays appears to be possible. As for the depth–selective

filtering characteristic to be applied, a synthetic DOF which simulates imaging by a lens with a aperture diameter around $P = 5.0$ mm seems to be optimum under our experimental conditions. This finding is above the average human pupil diameter to be expected for the illumination condition of our experiment (about 3.2 mm according to Ref. 12, pp. 51). Thus, a slightly smaller DOF than experienced in natural vision appears to be preferred in stereoscopic image representations.

In order to realize a true DOI display where a viewer can freely chose and change his or her point of fixation, a technique enabling precise real–time measurements of the viewer's lines of sight is needed. A video–based approach exploiting corneal reflections is currently under development at our institute. In a follow–up experiment in which the subjects will be allowed to randomly fixate points in stereoscopic space, we plan to re–examine the optimum depth–selective filtering characteristic found in the present exploratory study. Future work will also include consideration of more advanced post–filtering techniques for DOF simulation (e.g., the ray distribution approach recently proposed by Shinya[13]) which may also include simulation of lens diffraction effects.

Acknowledgments

References

[1] D.F. McAllister (Ed.): Stereo Computer Graphics and Other True 3D Technologies (Princeton University Press, Princeton, 1993).

[2] Y.-Y. Yeh and L. D. Silverstein, "Limits of fusion and depth judgement in stereoscopic color displays," Human Factors 32(1), 45-60 (1990)

[3] S. Pastoor, "Human factors of 3D imaging: Results of recent research at Heinrich–Hertz–Institut Berlin," Proc. of 2nd Int'l Display Workshop, Hamamatsu, Japan, pp. 69-72 (October 1995).

[4] A. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," Proc. SPIE 1915, 36-48 (1993).

[5] N. Hiruma and T. Fukuda, "Accommodation response to binocular stereoscopic TV images and their viewing conditions," SMPTE J. 102(12), 1137-1144 (1993).

[6] M. Wöpking, "Viewing comfort with stereoscopic pictures: An experimental study on subjective effects of disparity magnitude and depth of focus," Journal of the SID 3(3), 101-103 (December 1995). Also in: Deliverable No. R2045/HHI/AT/DS/C/014/b1, RACE–DISTIMA, 1992.

[7] M. Potmesil and I. Chakravarty, "A lens and aperture camera model for synthetic image generation," Computer Graphics 15(3), 297-305 (August 1981).

[8] S. F. Ray, Applied Photographic Optics (Focal Press, London, 1988).

[9] ITU-R, Draft Recommendation BT.500–7, Methodology for the Subjective Assessment of the Quality of Television Pictures (ITU-R Document 11/1030, 1995).

[10] H. A. David, The Method of Paired Comparisons (Charles Griffin & Co., London,1988).

[11] N. H. Mackworth and I. T. Kaplan, "Visual acuity when eyes are pursuing moving targets," Science 136, 387-388 (1962).

[12] Y. LeGrand, "Measurement of the visual stimulus." In: E. C. Carterette and M. P. Friedman (Eds.), Handbook of Perception, Vol.V: Seeing (Academic Press, New York, 1975).

[13] M. Shinya, "Post–filtering for depth of field simulation with ray distribution buffer," Proc. of Graphics Interface '94, Banff, Canada, pp. 59-66 (May 1994).
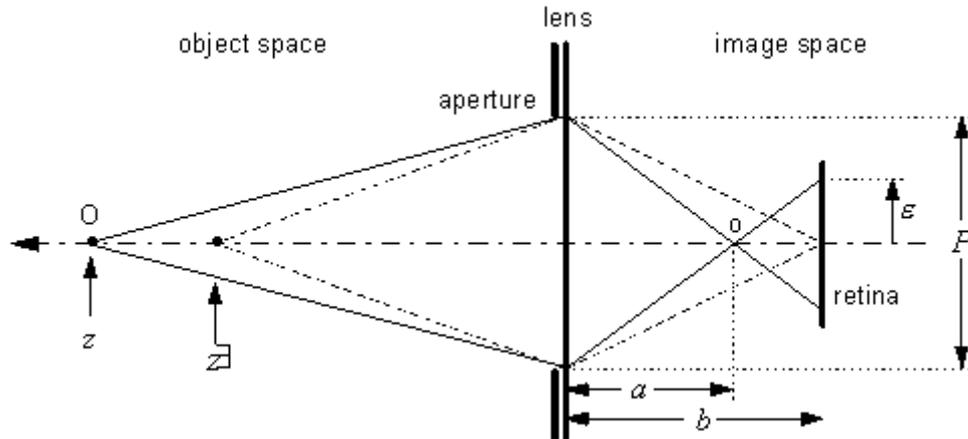
Figure 1: *Thin lens imaging geometry ($\bar{z}$: accommodation/fixation distance; P: diameter of aperture; b: image distance; a: distance of optically focussed image of O; e : radius of circle of confusion formed by O at the retina).*
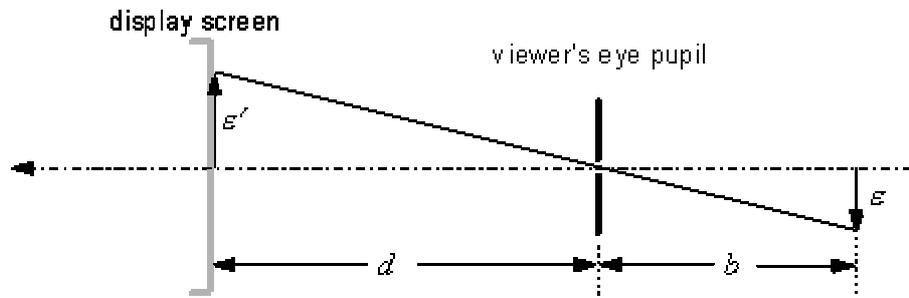

Figure 2: *Projection of e onto the display plane (e : radius of circle of confusion to be simulated at the retina; e ': radius of circular point expansion within the display image; d: distance between viewer's pupil and display).*
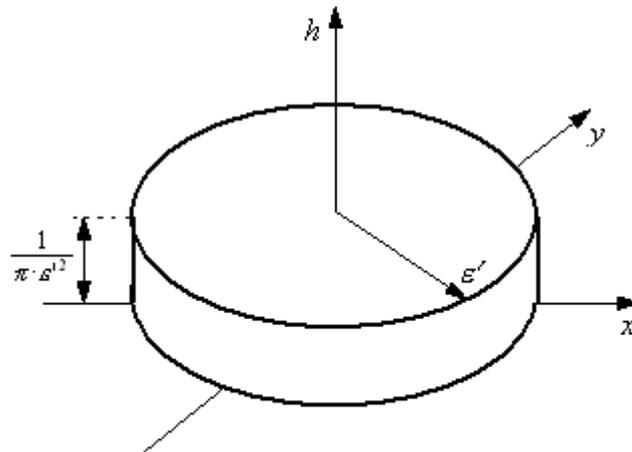
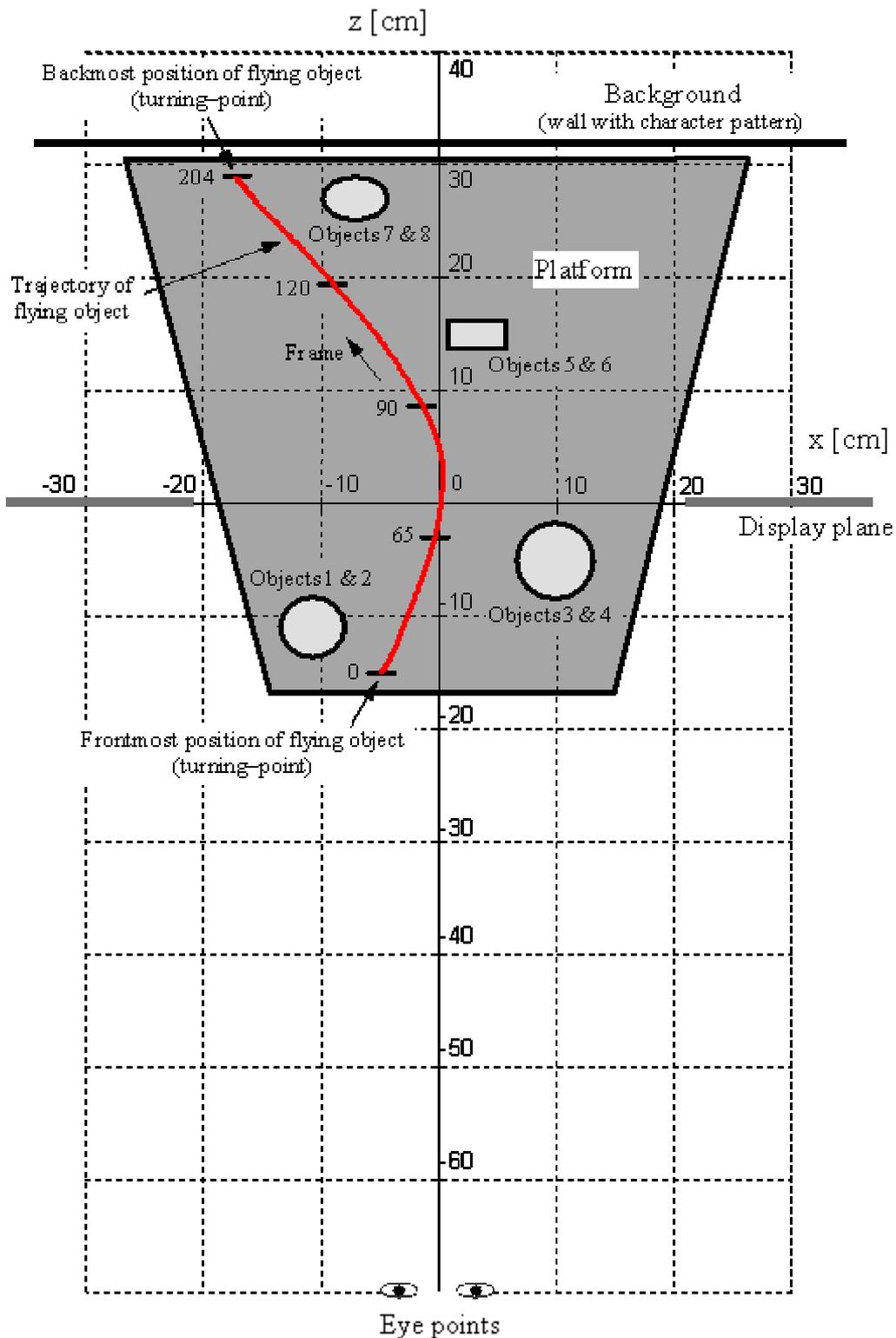***Figure 3:*** *Cylindrical point spread function h(x,* y, e ') (e ': radius of circular point expansion which varies with pixel position).
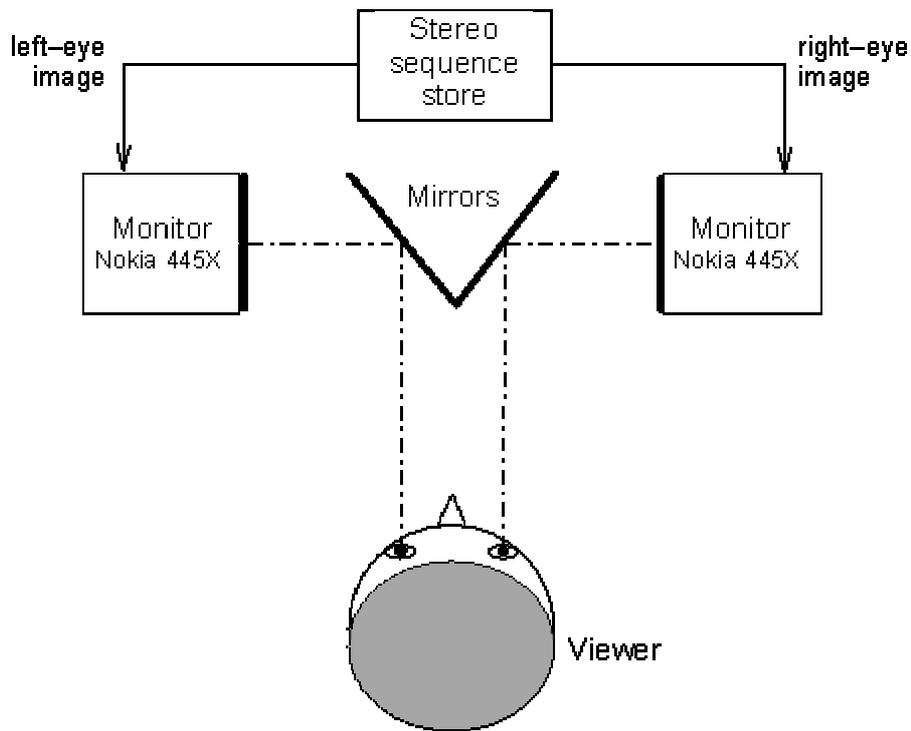
***Figure 4:*** *Two–mirror stereoscope set–up.*
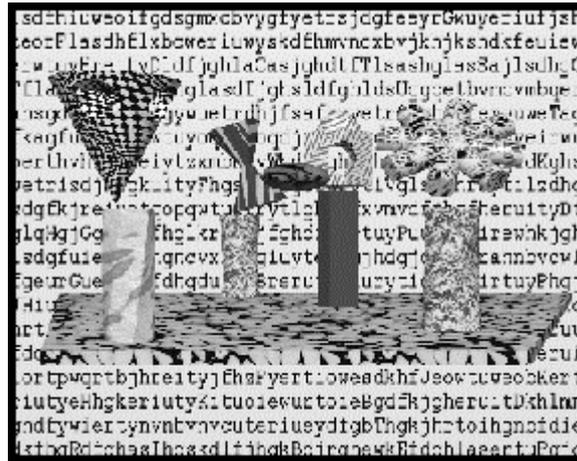
Figure 5*: Frame 65 of the original motion test sequence (right–eye image; the position of the flying object is approx. 3 cm in front of display plane).*

Figure 6: Schematic top view of computer–simulated 3D test scene. The flying object moves back and forth the plotted trajectory.Its velocity varies smoothly along this path (maximum near Object 5: $v_{x_{max}} \approx 0.11$ m/sec ; $v_{z_{max}} \approx 0.18$ m/sec).
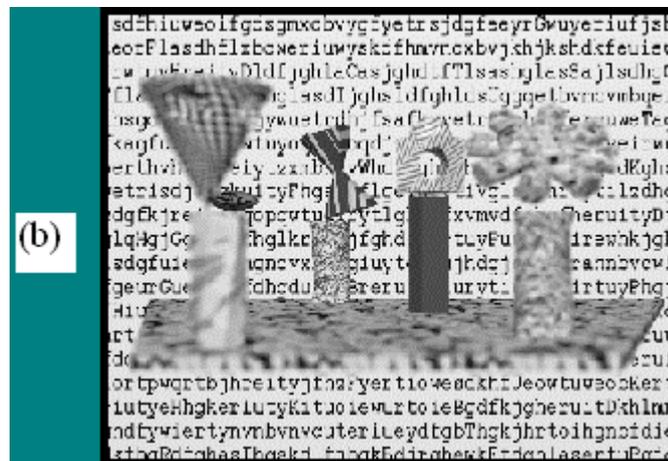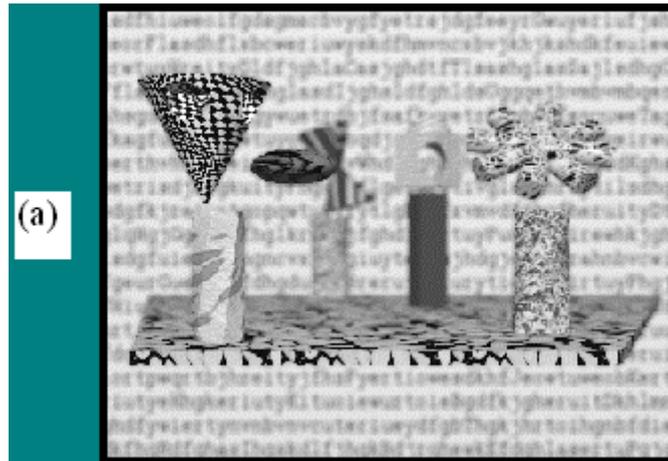
Figure 7: *DOF–filtered right–eye images (aperture diameter P = 10.0 mm). (a) Static image "front position" (flying object at its frontmost position, approx. 15 cm in front of display plane; identical with 1st frame of the motion test sequence); (b) Static image "back position" (flying object at its backmost position, approx. 29 cm behind display plane; identical with 204th frame of the motion test sequence).*
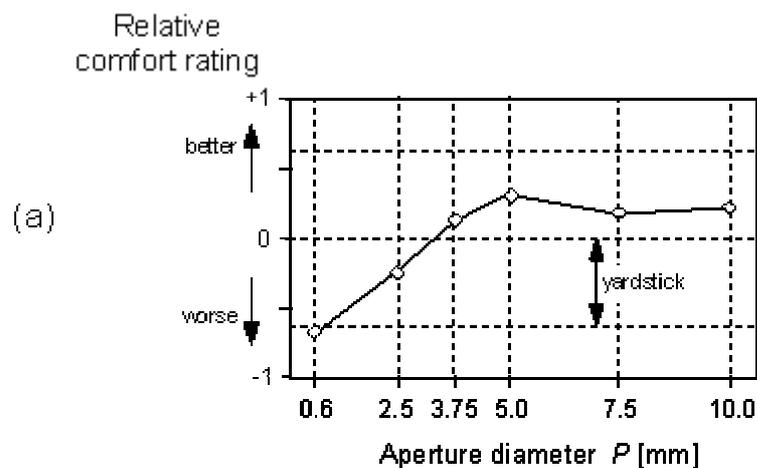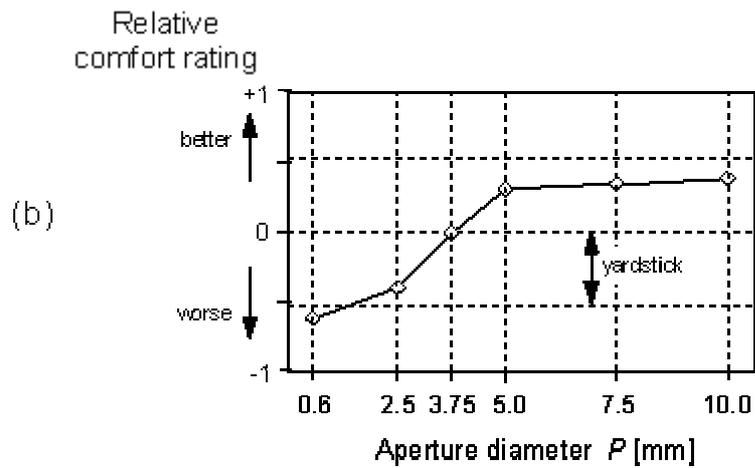
Figure 8: *Subjective ratings of viewing comfort experienced when viewing static stereoscopic images with six levels of DOF (statistically least significant differences are indicated by "yardsticks"). (a) Static image "front position". (b) Static image "back position".*
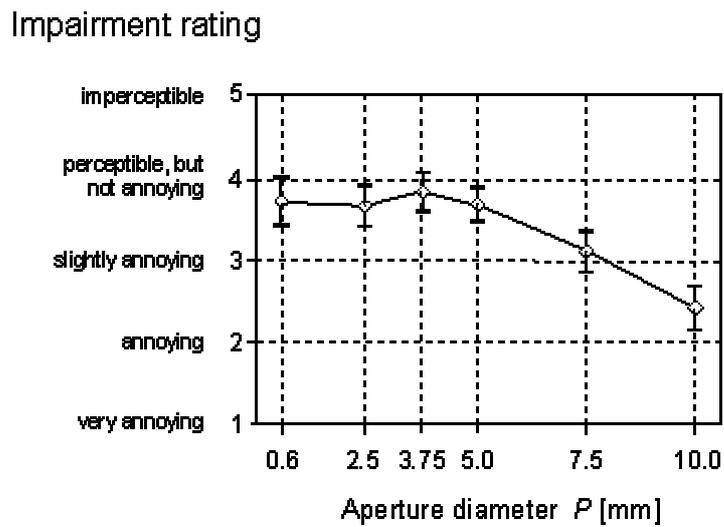
Figure 9: *Subjective ratings of viewing comfort impairment experienced when viewing the stereoscopic motion sequence with six levels of DOF.*

References

[1] A. Arditi, "Binocular vision", in K.E. Boff, L. Kaufman, I.P. Thomas (Eds.) *Handbook of Perception an Human Performance,* Wiley, New York, 1986.

[2] C. Ware and G. Franck, "Evaluating stereo and motion cues for visualizing information nets in three dimensions", *University of New Brunswick Technical Report TR94-082,* February 1994.

[3] S. K. Card, G. G. Robertson, J. D. Mackinley, "The information visualizer, an information workspace", *SIG CHI '91 Conference Proceedings,* pp. 181ff., ACM Press, 1991.

[4] G. G. Robertson, J. D. Mackinley, S. K. Card, "Cone Trees: Animated 3D visualizations of hierarchical information", *SIG CHI '91 Conference Proceedings,* pp. 189-194, ACM Press, 1991.

[5] R. A. Reaux, J. M. Carroll, "Human factors in information access of distributed systems", in G. Salvendy: *Handbook of Human Factors and Ergonomics*, 2nd edition, Wiley, New York, 1997.

[6] R. Skerjanc and S. Pastoor, "New generation of 3-D desktop computer interfaces", in: *Stereoscopic displays and virtual reality systems, Proc. IS&T/SPIE EI'97 Conference*, San Jose, Feb. 8-14, 1997.

[7] B. Shneiderman, "Designing the user interface – Strategies for effective human-computer interaction. 2nd Edition. Reading (MA) etc. : Addison-Wesley, 1992.

[8] G. R. McMillan, R. G. Eggleston, T. R. Anderson, "Nonconventional controls", in G. Salvendy: *Handbook of Human Factors and Ergonomics* , 2nd edition, Wiley, New York, 1997.

[9] J. Nielsen, "Noncommand user interfaces", *Comm. of the ACM*, 36(4), pp. 83-99, 1993.

[10] A.J. Glenstrup and T. Engell-Nielsen, "Eye controlled media: Present and future state", Thesis, University of Copenhagen, Denmark, 1995.

[11] R.J.K. Jacob, "Eye movement-based human-computer interaction techniques: toward non-command interfaces", *Advances in Human-Computer Interaction*, Vol. 4, Ablex Publishing Co., Norwood, N.J., 1993.

[12] S. Pastoor and R. Skerjanc, " Autostereoscopic user-computer interface with visually controlled interactions", *Digest of Technical Papers, SID'97 International Symposium*, pp. 277-280, Boston, 1997.

[13] M. Wöpking, "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus", *Journal of the SID*, 3(3), pp. 101-103, 1995.

[14] W. Blohm, I.P. Beldie, K. Schenke, K. Fazel and S. Pastoor, "Stereoscopic image representation with synthetic depth of field", *Journal of the SID,* 5/3, pp. 307-313, 1997.

[15] V. I. Pavlovic, R. Sharma, T. S. Huang, "Visual interpretation of hand gestures for human computer interaction: A review", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 19, No. 7, pp 677-695, 1997.

[16] S. Pastoor and M. Wöpking, "3-D Displays: A review of current technologies", *DISPLAYS*, 17, pp. 100-110, 1997.

[17]   R. Börner, B. Duckstein, O. Machui, R. Röder, T. Sinnig and T. Sikora, "A family of single-user autostereoscopic displays with head-tracking capapilities", submitted to *IEEE Transactions on Multimedia*, 1998.

[18]   G. Boerger and S. Pastoor, "Autostereoskopisches Bildwiedergabegerät mit natürlicher Kopplung von Akkommodationsentfernung und Konvergenzentfernung und adaptiver Schärfentiefe (Autostereoscopic display providing a natural link of accommodation distance and convergence distance and adaptive depth-of-focus)" DE 195 37 499, Patent Pending (1997).

[19]   L.-P. Bala, K. Talmi and J. Liu, "Automatic detection and tracking of faces and facial features in video sequences", *Picture Coding Symposium 1997*, Berlin, Germany, 10-12 September 1997.

[20]   R. Chellappa, C.L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey." *Proc. of the IEEE*, 1995.

[21]   H. Sako and A.V.W. Smith: Real-Time Facial Expression Recognition Based on Features' Positions and Dimensions, *Proc. of the ICPR,* 1996.

[22]   N. Oliver and A. Pentland, "LAFTER: Lips and face real time tracker. *IEEE Conference on Computer Vision and Pattern Recognition* , San Juan, Puerto Rico, 17-19 June 1997.

[23]   L. Young and D. Sheena: "Methods & Designs: Survey of Eye Movement Recording Methods". In: *Behavior Research Methods & Instrumentation*. Vol. 7(5), 1975, s. 397-429.

[24]   K. Talmi and J. Liu, "Eye and gaze tracking for visually controlled interactive stereoscopic displays", *Image Communications,* 1998 (in print).

[25]   J. Liu, "Determination of the point of fixation in a head-fixed coordinate system", *14th International Conference on Pattern Recognition,* Brisbane, 16-20 August 1998.

[26]   S. Nagata, "How to reinforce perception of depth in single two-dimensional pictures", *Proceedings of the SID* , Vol. 25, No. 3, pp 239-246, 1984.