COMPARISON OF DIFFERENT 3D (STEREO) VISUALIZATION METHODS – EXPERIMENTAL STUDY

László Vajta[1], Tamás Urbancsek[1], Ferenc Vajda[1], Tamás Juhász [1,2]

[1] Department of Control Engineering and Information Technology, Budapest University of Technology and Economics, Magyar tudósok krt. 2, 1117 Budapest, Hungary
[2] Department Virtual Engineering, Fraunhofer Institute for Factory Operation and Automation, Sandtorstrasse 22, 39106 Magdeburg, Germany

urbi@iit.bme.hu (Tamás Urbancsek)

Abstract

3D (stereo) visualization is a fast developing topic in mechatronic simulation, design, and teleoperation. Plenty of methods are in the experimental stage.

Although the stereo techniques using anaglyph, liquid crystal shutter or polarizer glasses, or the auto-stereoscopic lenticular displays are directly stimulating the binocular human cues, they can also have characteristic influence on the monocular ones. We would like to emphasize the importance of motion parallax phenomenon, which creates an illusion of depth when the observer has a relative motion to the sight, and the location / coverage of perceived points is changing according to this motion. The resulting effect can be different and it is not a trivial task to choose the optimal solution for a given problem.

The main goal of our research was a systematic comparison of the subjective perception of depth delivered by the aforementioned 3D techniques extended with dynamic motion. We developed an experimental cell, where we projected a given show with various methods for a group of voluntary people without informing them about the actual technology. We asked the test persons to evaluate questionnaires looking at various static or dynamic 3D scenarios presented in the cell using these techniques, concerning their subjective impressions on their spatial sight while either wearing anaglyph / LCS / polarizer glasses or not.

In this article we report the partly surprising results, that visual depth sensation delivered by motion parallax technique compares to the efficiency of traditional 3D visualization techniques. The achieved results can be integrated in planning of simulation and teleoperation systems.

Keywords: 3D, Stereo, Visualization, Teleoperation, Experiment

Presenting Author's biography

Tamás Urbancsek is a research assistant at Department of Control Engineering and Information Technology, Budapest University of Technology and Economics since 2004. His interests and main research fields include telerobotics, multi-agent robotics, image processing and virtual reality. He has published more than 15 lectured articles in these research areas.

ISBN 978-3-901608-32-2

1 Introduction

Mobile platform teleoperation means: operating a mobile platform at a distance. This solution is frequently used in difficult to reach environments: to reduce mission cost, and to avoid loss of life. Although some restrict the term teleoperation to denote only direct control (i.e., no autonomy), we consider teleoperation to encompass the broader spectrum from manual to supervisory control [1]. Furthermore, the type of control can vary and may be shared / traded between operator and vehicle.

The human perception of depth is a complex sensation. It has at least three different components: the so called extra-retinal (not discussed here), the monocular and the binocular ones. As the most teleoperated mobile platforms have visual guidance, we can distinguish them along this criterion.

Most of the teleoperated systems work by use of monocular image transmission. If the field of view is defined by motion tracking of the operator's head, the visualization equipment is normally a head mounted display. Plenty of experiments proved, that the head position feedback is not essential [2]. Although tracked systems produce more realistic feeling, and therefore their application is a must in simulator systems, their realization is expensive. We can spare their installation, if the personal is well skilled. On the other hand the need for stereoscopic vision is a more challenging problem.

The monocular impression is the mixture of more, well-known phenomena as the perspective, shadowing, atmospheric distortion, a priori expected size of the objects, texture distortion, etc. Even in case of one-eye (monocular) sensing, the human brain interprets some kind of depth information. Although the capability of the monocular perception of depth is limited and it is sometimes inaccurate, it has essential role in the global sensing procedure.

Binocular sensation (frequently called as stereo vision) is the most trivial component of the depth sensing. Due to the 4-6 cm separation between the eyes, each eye has a slightly different viewpoint. The images from two different perspectives are sent to the brain and their difference – which is termed parallax – is interpreted as depth.

The artificially generated visual impression is an inherent part of the teleoperation interface. The use of visualization methods in this field, especially in case of remote controlling of mobile systems (teleoperation) needs the possible maximum level of reality. There are plenty of displaying methods under development and in use for this purpose [3].

We may not forget that monocular feeling plays an important role in our visual sensation. (Our experiments proved that by using monocular sensing only – for example the transmitted images of the onboard camera – we can even drive the mobile platform remotely with high safety).

The stereo depth sense is produced through evaluation of the horizontal shift of corresponding points on two images. Even in case of random dot structure the human brain can find the corresponding pairs and represent them as in depth distributed information. But there are other image features, which are heavily influenced by the depth, too. In our understanding, the key issue for the monocular depth sensing phenomenon is the depth sense from the motion, and the reflection.

The projected image of a true 3D scene, which contains objects in different distances, will change the arrangement of the objects on the image in case of any change in the viewer's position. Objects on the image are covering each other, whereas this coverage varies according to the viewing angle. Continuous movement of the viewer causes continuous change in this coverage. By displaying the image of the moving camera on a monocular screen impressive depths feeling can be achieved. Actually this is the case if a vision guided robot moves along its trajectory.

A very important question arises. Is the binocular stimulation of the human vision system a must for nearly realistic impression? The main goal of our research is to find an adequate answer on this question.

There are techniques that provide stereoscopic vision without the need for specialized viewing equipment: these are exploited in the autostereoscopic displays. It is incontrovertible, that the use of a special auto-stereoscopic display is the most advanced solution (to set aside its price), because such displays provide three-dimensional stereo images over a range of viewing angles without the need for special viewing glasses (or a head mounted unit, which definitely needs the head movement tracking).

A lenticular display consists of a horizontal array of cylindrical lenses placed in front of interleaved pictures of a given object from slightly different viewing angles. The device is designed so that at a given viewing angle, only one of the set of interleaved images can be seen. It has a special shape crystal layer over the surface of a simple LCD display. These crystals distribute each pixel into exactly one of the several (e.g. 8) different viewing directions.

The solution is sensitive on the movement of the viewer. There are displays on the market, which are able to track the movement of the user's eye, and modify the displayed images accordingly. This solution produces true depth sense, and allows relative free movement of the user – without the need for glasses. Portable auto-stereoscopic 3D imaging becomes popular for example on laptop computers.

Fig. 1 illustrates how a lenticular display creates this effect based on light refraction:

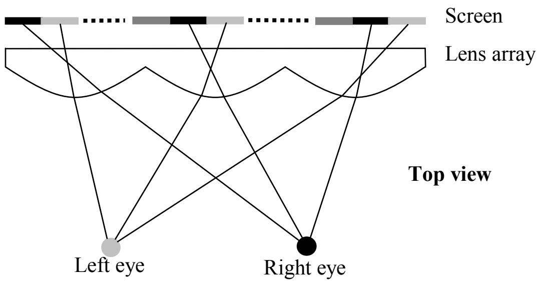

Fig. 1: The autostereoscopic lenticular display

However, the achievable resolution of such a device is limited by large files sizes, and the available monitor resolution due to the need for interleaved images. Another drawback is that despite its incorporation of many depth cues, it is still a two dimensional representation, and it is incapable of exploiting accommodation. Only truly spatial three-dimensional displays will be able to provide this.

Note that in fact all of the three-dimensional imaging techniques described up to this point has been two-dimensional representations with enhanced depth cues to create the illusion of a third spatial dimension. However, no matter how complex they are, they are still two-dimensional and ultimately flawed. This is because they lack the ability to appease the eyes need for ocular accommodation in experiencing a three-dimensional world. Spatial three-dimensional displays will allow the eye to focus at different depths throughout the display and truly experience all three dimensions. The volumetric display technique is still in its infancy, but it is worth further investigation because it is rich not only in fundamental science from different fields but also in skilful, detailed engineering.

2 Stereoscopic systems

Stereoscopic systems need two separate images from the same scene. The human eyes must sense their respective images only. We can classify the stereo displays based on the method of the separation:

- passive separation
- active separation

Every solution has its own inherent limits, which can be more or less disturbing and the level of disturbance depends on the individuals themselves.

2.1 Passive separation

Stereoscopic systems, which use *passive spectacles*, separate the two – in space overlapped – images for left and right eyes by some kind of passive filters. Wearing such extra viewing equipment can be a bit inconvenient (especially for people having normal eyeglasses), since an object has to be placed on the head. Nevertheless these are simpler and can offer more “real” sensation then the ones without glasses. In our experiments we used various techniques that need wearing spectacles, which are discussed also in [4].

Anaglyphs use spectral separation, where the used spectacles are having colored optical filters (red-cyan glasses, see Fig. 2) in fact [5]. The method can be described as follows. The left lens of the glasses works as a color-filter (usually red), while the right one has its complement color (usually cyan, represented by green and blue pixels on displays). So, two images are separated for the two eyes in their color spectra. The human brain is capable of putting them together into a color 3D image.

Fig. 2: Anaglyph glasses

This is the simplest 3D visualization method; however the better the filters are, the better quality of perception can be achieved. There are several low quality products on the market that are cheap, but give poor results: they usually don’t match the display’s color range or have big cross-talk between the color filters for the eyes).

Note that there are other color-based 3D technologies with glasses, such as Chromatek or Pulfrich. Using Chromatek goggles the colors of lower wavelength (blue) are perceived like those objects would be farther, while objects with colors of higher wavelength (red) seem to be closer. Darker colors are perceived later than brighter colors, which theory is used by Pulfrich glasses, while the original image is moved. There is a special usage area of these two methods, so we did not use them in our experiments.

Other solutions are using *polarized lights*, where the two images have different polarization, which is aligned with the polarization filters of the viewing glasses. There are two methods of polarization-based 3D visualization: it can use either active or passive displays. In passive methods the two images of the 3D view are projected onto the same canvas with differently polarized light beams (where polarization can be either linear or circular). It is important, that the canvas must not change the polarization of the light. In linear case the polarization of the two projections can be perpendicular to each other (or can be opposed using circular polarization), so using the respective polarizing filters again in the glasses, only a single image will be visible for each eye at a time. Simple projectors can be used with polarizing filters for this purpose.

In active methods, a display creates natively the two sorts of polarized light: this is nowadays only possible using special LEDs (put into LED matrix displays). The same polarized goggles need to be used to view such displays.

2.2 Active separation

Stereoscopic systems with *active glasses* usually use time-overlapped images. Using *LCS* (Liquid Crystal Shutter) glasses, the two images are displayed in a time-sharing way. These spectacles are programmed shutters, which direct the images intermittently to the right and left eyes. In a frame one of the lenses is opaquely black, while the other is transparent: thus the right eye can see only the right image (see Fig. 3). Each frame the opaque / transparent status of the lenses is switched. The biggest disadvantage of this method is that the display frequency must be duplicated (in fact, this can still be not enough); otherwise, blinking can be very disturbing. Since nowadays displays (TFT) and projectors (DLP) cannot be set to higher frequencies, this method has also its own limits. Setting transparency is simply solved by a single liquid-crystal cell for each lens, which can be either black – if electrical potential is connected to the electrodes – or transparent if there is no voltage. This is synchronized with the continuous refreshing of image frames.

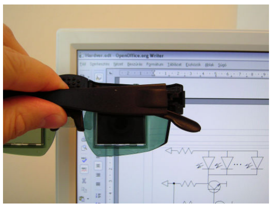

Fig. 3: LCS glasses

This active stereo method is also a jitter-sensitive solution – the smallest time-shift between the image display and the glass control “kills” the stereo feeling.

There are other glasses-based solutions as well (e.g.: Infitec [6]), but we used the above ones (anaglyph, LCS and polarizer glasses) in our experiments.

3 Hardware for the experiments

In order to have sufficient 3D sensation, installing special hardware components is essential. In recent years, several solutions have appeared on the market of 3D devices, so 3D virtual reality is also available in home environments.

In the course of our research we also tried to find solutions for 3D visualization at larger scales (i.e.: LED displays with diameter of several meters); however, in the psychophysical experiments this article is about, we did not use them, so we confine now in this section to present the equipment used in these very experiments.

During our experiments (see section 5) we used mostly goggles-based technologies, partially combined with motion parallax. In the following section we describe the test environment of the experiments.

3.1 Environment

We had furnished a small room for the experiments. The room contained a table with a chair and the accessories of the experiments. At a particular time, only one participant and the experiment leader stayed in the room. The specification of hardware accessories was the following:

- PC: AMD Athlon64 Venice 3000+ CPU, 1024 MB memory, Inno3D Geforce 6600, PCI-E, 256 MB video card
- Projector: Toshiba DLP (two pieces), 2 circular polarizing filter one per each (one polarizing clockwise and one counter-clockwise)
- Polar glasses (also one CW and one CCW)
- LCS glasses with Infra Synchronizer
- Anaglyph glasses
- Polarization preserving canvas

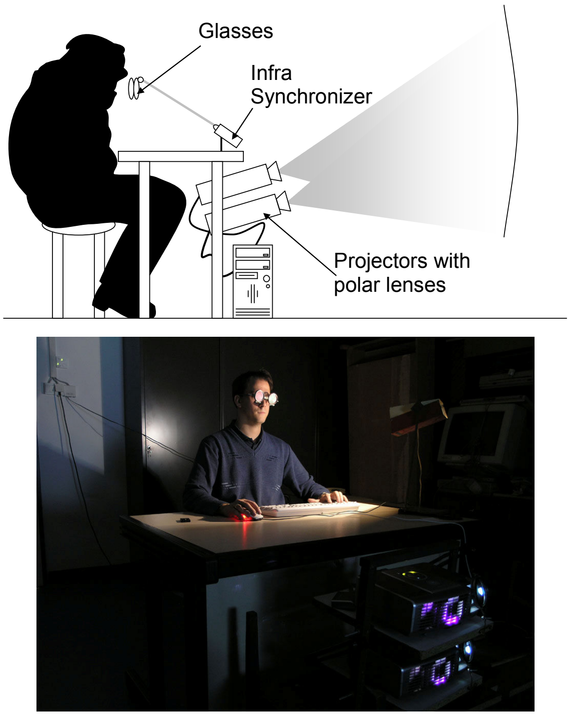

Fig. 4: Environment of the experiments

In order to carry out the experiments we had to solve some simple issues. These are described in the followings.

3.1.1 Polarization preserving canvas

Polarization is the property of light (or other electromagnetic waves) that describes the direction of the transverse electric field. Thus, polarization is the direction of oscillation in the plane perpendicular to the direction of travel. Polarization can be linear, circular or elliptical. If a linear polarized filter is in the way of a perpendicular linear polarized light, no light is transmitted. The same happens if a circular polarized filter is in the way of a circular polarized light of the opposite direction. Therefore, two different lights can be transmitted to the two eyes from the same direction.

Both projectors have been installed with a circular polarized filter, so two images with different polarization are projected to the canvas. Trough the polar filter glasses, both eyes can see only one image.

Most surfaces modify the polarization of the incident light (usually by scattering). However, in our application this would cause an unusable result. There are polarization preserving canvases on the market, but they are rather expensive. By examining some inexpensive surfaces we found that cardboards coated with silvery paint retain polarization sufficiently, so we prepared the reflective surface using this kind of cardboards.

3.1.2 LCS synchronization

In the course of our project we had to create some additional hardware and software components to fulfill the qualification test. The LCS glasses we used are light controlled wireless active devices. They were designed for using with PAL or NTSC systems, so the shutters can be switched at appropriate frequencies (50Hz or 59.94Hz). Every video card is capable of setting a refresh rate of 60Hz, so NTSC mode could be a good choice, only we had to synchronize with the VS signal of the VGA connector. The infra protocol of the original synchronizer was measured by an ad hoc hardware, and a simple microcontroller based device was built (see Fig. 5) to mimic the original behavior, and transmit signals to synchronize the switching of the shutters of the glasses with the projector's vertical refresh signal.

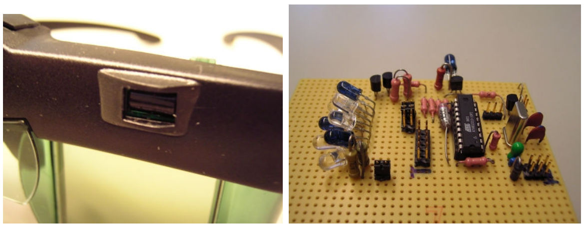

Fig. 5: Infra Synchronizer

4 Software for the experiments

At the Mobile- and Microrobotics Laboratory of our department we have developed a software that supports intuitive comparison of various 3D visualization methods. We asked a group of voluntary people to evaluate questionnaires, concerning their impressions on spatial sight while looking at various 3D scenarios on the canvas presented by this program.

The software is written entirely in C# and is based on the managed version of Microsoft's Direct3D SDK (version 9.0c, later referred to as D3D). We use various pre-generated .X scenes and advanced (HLSL) shader techniques for visualization.

4.1 Structure

The application has a main module that presents a given 3D scenario by means of various visualization techniques. Using the user input module, a range of parameters of the visualization (the technique itself, the position and orientation of virtual cameras, etc.) can be changed interactively. This component is in close connection with a script interpreter module that allows automated control of some of these parameters, and monitors the user's behavior making the anticipated objective evaluation later possible (for the evaluation procedure using Matlab see section 5).

4.2 Implementation of 3D methods

The program offers four well-known 3D visualization techniques: the motion-parallax (when only monocular cues are presented), the anaglyph (images for left and right eyes are separated in spectral domain), the liquid crystal shutter (images are separated in time domain) and the dual-view methods (image separation using polarization).

We use the (C-like) HLSL high level shading language [7] to implement our 3D techniques. The so called "shaders" are programs that run on the specific parts of the GPU (graphics processing unit) instead of the main processor. The shader codes are automatically compiled into an efficient intermediate effect language (and later to machine code of the actual GPU) in runtime. Basically a vertex program is a graphics function used to perform mathematical operations (for example perspective projection to screen space) on the objects' vertex data (position, normal, etc.). The pixel- (also known as fragment-) program is a graphics function that calculates effects on a per-pixel basis after the output of the vertex shader (i.e. triangles) is rasterized. The basics of vertex- and pixel shader programming can be found in [8].

For binocular techniques we need to generate images for both eyes' perspective and make them separately stimulate the left and right eyes, respectively. According to the 3D technique being used at a given time, the composition of these images can appear as the final image, or we pick a single one to be displayed (for LCS glasses) or show both ones simultaneously (using two projectors and polarization lenses).

Although the scene being viewed at can be static, any camera movement can cause significant increase in the depth impression of the viewer. We allow an option to use a small amplitude harmonic function to automatically rotate the camera around its target point to enhance this depth cue.

4.2.1 The basic monocular method

In case of the *motion-parallax method* we can render the projection of the 3D scene directly to the screen in the usual way:

For each object in the scene we set up the transformation matrices and the material parameters, and then feed the mesh data of the dynamic vertex buffers through the programmable graphics pipeline using our shader routines in order to get the shaded result.

In order to avoid flicker, we draw everything first to a back-buffer image and after synchronization with the display hardware, this image can be presented on the screen (this is the well-known double buffer technique [9]).

The camera that is used in this mode will be referred later to as the cyclopean or central camera.

4.2.2 Binocular methods

For binocular methods we distinguish two camera models: the one with parallel viewing axes and the one with convergent eyes, looking at the same target point in the space (also known as "toe-in"). In our implementation there is no support for asymmetric frustums, for the latter case we use the method shown in the following Fig. (when the cameras are rotated towards each other around the world's vertical axis):

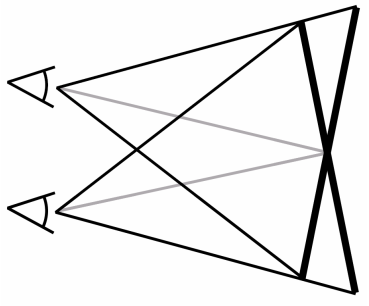

Fig. 6: The case of convergent cameras (Top view)

The left and right eyes have a horizontal offset (*base*), the half of which is the distance from the central (cyclopean) eye's origin for both eyes. We treat this value with a ± sign, so it is easy to switch the roles of the left and right eyes, when the current setup needed that (e.g.: by changing the order of the projectors dual projection mode using polarizer filters).

For the *convergent case* the left and right cameras are transformed from the central one the following way (the camera's **up** axis is the unit vector that is vertical in camera space, pointing upwards):

$$\mathbf{z} = \mathbf{eye}_{center} - \mathbf{target}_{center} \tag{1}$$

$$\mathbf{offset} = \frac{base}{2}\mathbf{up}_{center} \times \frac{\mathbf{z}}{|\mathbf{z}|} \tag{2}$$

$$\mathbf{eye}_{L,R} = \mathbf{eye}_{center} \mp \mathbf{offset} \tag{3}$$

$$\mathbf{target}_{L} = \mathbf{target}_{R} = \mathbf{target}_{center} \tag{4}$$

If we change the sign of the **offset** vector, we can switch the roles of the left and right cameras. Additionally, the *parallel case* implies that the camera targets have to be relocated, too:

$$\mathbf{target}_{L,R} = \mathbf{target}_{center} \mp \mathbf{offset} \tag{5}$$

For the *anaglyph method* we have to generate a composite picture, as this one separates the left and right source images only in spectral domain (i.e. with color-filters). In our application for each frame we render the left and right images into dynamic textures, which have the required resolution that matches our display. In D3D these dynamic textures can be stored in the fast video memory. After setting up these new rendering targets, and adjusting the virtual camera to the view of the given eye, we can render 1-1 image into those textures using the usual way (through the programmable pixel pipeline using shaders). As we would like to composite the result image, finally we need to render a single rectangle (i.e. two triangles) that fills exactly the entire screen, thus we can calculate the color of each pixel of the result image using a simple and fast pixel shader.

When c_1 and c_2 are the source pixel colors (in RGBA order) at the current position, we use the following formula in the compositor pixel shader to calculate the color of the current pixel (Note that we neglect transparency here):

$$c = \begin{bmatrix} c_{1R} & c_{2G} & c_{2B} & 1 \end{bmatrix} \tag{6}$$

When we use the *LCS method* we render the image only from one eye's perspective for a given frame. As we can directly render to the screen (like we did in the monocular case), this implies that this technique requires much less resource than the anaglyph one.

The LCS glasses have a control hardware using the "vertical sync" signal of the graphics card's video output. This synchronizes the switching between transparent and opaque states of the lenses. We use this fact in our software, and tell the D3D device to swap our double-buffers right at the same time.

With this technique each time a frame is rendered, in the next frame the other eye's view comes into consideration. We make the user also possible to switch between the L-R and R-L cases with a keyboard hit, as we are unable to predict which shutter is the opaque one, while we present the image destined for the left eye, for example.

The last binocular technique is using the *circular polarized lenses*. As we have two projectors equipped with polarized filters displaying their own image on the same canvas, this method requires each frame rendering the scene twice, from both eyes' perspective. We have to present these images simultaneously on the projectors connected to the two video outputs of the graphics card. For this purpose we have to use the dual monitor configuration in the display driver.

We encountered a problem that is in connection with the D3D API being used in our application. Namely, it is not possible to have a rendering device operating on more than one display in fullscreen mode. In the managed D3D API only a single fullscreen device can exist at a time: the one that can own the graphics hardware exclusively. Each time we want to use the polarizer glasses technique, our solution is to switch back to windowed mode, and create a rendering window, which has a height value that matches the native vertical resolution (*H*), but has a width value that is two times the native horizontal resolution (*W*) of a single display (i.e.: *2560 x 1024* pixels when using *1280 x 1024* resolution on the projectors). First we define a viewport with upper-left corner at (*0; 0*) with dimensions *W x H*. We render the scene to that viewport the usual way, and then relocate the viewport to position (*H; 0*) while its size remains *W x H*. From the other eye's perspective we render the scene once again into the new viewport's back buffer. As the *2W x H* big window stretches exactly over the two displays, each half (a viewport's image) will finally be shown on its representative monitor (projector).

Finally we have to note that in windowed mode no fixed framerate can be guaranteed, thus this is not applicable for the shutter method (that needs exclusive fullscreen usage of the graphics hardware, because the goggles are synchronized with the screen refresh signal). The anaglyph and motion-parallax methods in our application run always in full-screen mode, but this is only a convenience.

4.3 Supporting psychophysical experiments

We let a group of persons evaluate a series of 3D scenarios in the aspect of 3D spatial impressions. During these psychophysical experimental tests the subsequent user input was disallowed and some of the scenario-parameters (3D scene, visualization technique, monocular camera / binocular camera with eye-base distance) were controlled by the script interpreter module of our application.

During each experiment a pre-generated text file is being read line by line to generate new 3D scenarios that will be presented sequentially to the user. He / she might be asked to take either the anaglyph (red-cyan lenses), the LCS or the polarizer goggles on or off, before the new setting is shown.

The user is asked to give scores for each setup, and these values will be stored by the software in a file, which makes evaluating possible, after all experiments have finished. We know that sometimes simply the fact the user has to wear extra goggles can also be a negative factor for subjective judgment. Thus we generate also scenarios, where the left and right virtual cameras are coincident, and we still ask the user to wear the goggles, in order to analyze this effect.

5 The experiments

This chapter describes the design and the result of a series of experiments we conducted with 44 participants. The subjects were students of Computer Science.

The goal of the experiments was to quantize the psychovisual perception of depth delivered by the aforementioned 3D techniques. While other researchers concentrated on the efficiency of 3D techniques in combination of the shadowing [11], or particular applications [12], this paper investigates the efficiency these techniques combined with the motion-parallax effect, which plays an important role in mobile robot teleoperation.

5.1 Measurement model

As depth perception is a subjective quantity, one can have the question whether it is measurable at all or the measurements delivered by two different scenes are really comparable. Hereafter we show that it is possible building an appropriate measurement model.

The first problem is that the attendants cannot quantize exactly their depth experiences even on an exactly defined scale, thus it can happen that one estimates the same experience differently at a later time. This effect is the measurement noise. We can suppose that it is a Gaussian noise with zero expected value.

A bigger issue is that the same scene generates different depth sensation level to the observer under the same circumstances but at a later time. The first impression with a novel technique may cause over-estimation of the experience. Therefore, the measured value depends on the preceding scenes viewed by the experimental person. We can model this effect with an internal state of the observer person.

In the following we presume that the depth perception caused by every scene has an intrinsic component, which is independent from the observer person (and from his internal mental state caused by previous scenes) and defined as the expected value of the measurement of that scene.

The following Fig. below shows the model of an experiment.

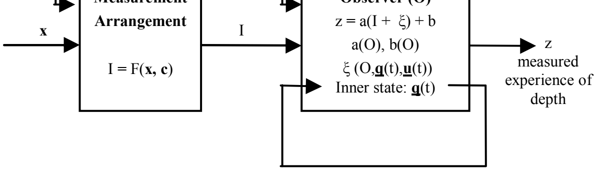

Fig. 7: Model of the experiments

- **x** denotes the parameters of the scene arrangement (which technique is used, is the virtual camera moving, depth cue)
- **c** denotes all the other parameters of the measurement arrangement, which affect the intrinsic depth perception level and we have to set them constant during the series of experiments.
- I is the intrinsic depth perception level. We want to estimate this value at the different scene arrangement parameter values (**x**).
- F is the function of **x** to I.
- The aggregated measurement noise (ξ) depends on the internal state of the observer (**q**) the environmental effects (**u**). Both components are time-varying (t).
- Every person has his internal habits a personal bias (b) and a personal interpretation of scale (a). We model these individual properties as a linear transformation of the intrinsic depth sensation.

5.2 Design of Experiments (DoE)

In this series of experiments our goal was to compare the intrinsic depth perception level of the different measurement arrangements by asking the observer person about it. We used two approaches in the questionnaire:

- we asked to report his / her opinion about the absolute level of quality the how the given technique delivers depth sensation and
- we asked how it compares to the previous arrangement.

5.2.1 Distortion effects and their elimination

Based the model above, we used the following methods to eliminate the distortions:

- We brought every observer person to the same state in the beginning of an experiment (**q**(0)) by giving them the same a-priori information about the experiments.
- We reduced every other outer effect (**u**) by performing the experiments in a silent dark room, where only the projected computer screen image was visible.
- We assured the independency between **x** and **q**.
- We tried to keep the parameters of the measurement arrangement (**c**) to the same level during the series of experiments.
- We chose the order of successively measurement arrangement parameters that the distortion effects can be eliminated with statistical methods

5.2.2 Measurement arrangement parameters and their order

The measurement arrangement parameter space can be split in three factors:

Applied 3D technique (X_1)

- Anaglyph technique (A)
- Shutter technique (S)
- Dual stereo technique (D)
- Monocular technique (M)

Binocular cue: (X2)

- No binocular cue (0)
- Binocular cue with convergent camera model (T)
- Binocular cue with parallel camera model (P)

Camera rotation (X3)

- Camera rotates (R)
- Camera does not rotate (N)

The following Fig. shows the domain of the parameter space and the adjacency of their possible values. We define two parameter values adjacent if they differ in a single factor.

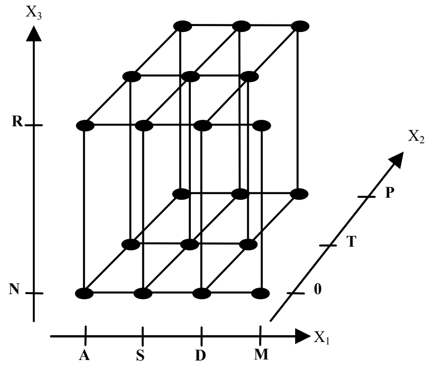

Fig. 8: Adjacency in the parameter space

By defining the order of the measurement arrangements (scenes) the following criteria had to meet:

- We wanted to estimate the intrinsic depth perception level in all of these points equally therefore we had to apply a full factorial design, thus, we showed every scene (node of the graph) once to every observer person.
- As the questionnaire asks to compare the currently viewed scene to the previous one it was recommended that the successive measurement arrangements differed in a single factor, thus they should be adjacent.
- We pointed out in 5.2.1 that **x** and **q** had to be independent. This means that every measurement arrangement (vertices in the parameter space graph) and also every adjacent scene pair (edge in the graph) had to take place randomly in the order of scenes presented for an observer.

Thus, we had to find Hamiltonian paths in the parameter space graph, which fulfill the criterion of independency defined in section 5.2.1.

We designed the whole series of experiments with our heuristic DoE algorithm written in MATLAB that found 18 long “almost Hamiltonian paths”, which fulfill the criterion independency.

The following table shows the design of experiments for the first 20 observer persons. A scene id denotes a node in the graph (a parameter of a measurement arrangement). The associated id is an ordinal number of the node interpreted as the number $X_1X_2X_3$ of variable number system: 0 denotes (A,0,N), 1 means (A,0,R) etc.

Table 1: Design of Experiments (for the first 20 subjects)

Nr	**Scene order in node id ($X_1X_2X_3$)**																			
1.	17	13	12	14	8	10	6	7	9	11	5	3	1	19	18	0	2	4	16	15
2.	15	13	19	1	5	4	0	6	12	16	10	8	14	2	3	9	7	11	17	18
3.	18	6	10	11	7	19	13	1	0	12	14	16	17	15	9	3	2	4	5	8
4.	8	9	15	14	12	18	19	7	1	3	5	11	10	6	0	4	16	17	13	2
5.	2	8	6	18	12	13	7	9	3	15	17	11	5	1	0	4	10	16	14	19
6.	19	7	6	8	2	0	18	12	16	4	5	17	15	13	1	3	9	11	10	14
7.	14	15	3	5	4	2	0	1	7	13	17	16	12	18	6	8	10	11	9	19
8.	19	18	0	6	10	4	2	14	15	17	16	12	13	7	11	5	3	9	8	1
9.	1	13	15	14	16	10	4	0	12	6	18	19	7	9	8	2	3	5	11	17
10.	17	5	1	7	6	0	18	19	13	15	9	11	10	8	14	2	4	16	12	3
11.	3	1	5	17	11	7	19	18	12	0	2	14	16	4	10	6	8	9	15	13
12.	13	17	5	11	9	7	1	3	2	8	6	12	18	0	4	10	16	14	15	19
13.	19	1	13	12	0	18	6	7	11	17	15	3	2	14	8	10	16	4	5	9
14.	9	8	2	0	1	19	13	7	6	10	11	17	5	3	15	14	12	16	4	18
15.	18	6	7	13	19	1	0	2	3	15	17	11	9	8	14	16	10	4	5	12
16.	12	6	18	0	1	5	4	2	8	9	7	19	13	17	11	10	16	14	15	3
17.	3	1	19	18	12	13	15	9	7	6	0	4	16	14	2	8	10	11	17	5
18.	5	1	7	13	19	18	0	6	12	14	8	9	15	17	16	10	4	2	3	11
19.	11	7	1	13	12	16	17	5	3	15	14	2	4	10	8	6	0	18	19	9
20.	9	3	1	13	19	7	11	5	17	16	12	6	18	0	2	14	8	10	4	15

The answers of the questions were a number between 1 and 9. An example of a presented scene was the Direct3D SDK’s well-known Dwarf model over a high-detailed textured terrain:

Fig. 9: A sample of a presented scene

5.3 Interpreting the results of the experiments

We performed these tests on 44 experimental persons. First we evaluated the answers for the first question, that is, what was the subjective sensation of depth delivered by the given scene.

5.3.1 Evaluation of function F

Every observer person saw the same set of measurement arrangements; therefore the average and the standard deviance of the answers should be the same. The average of all samples (answers) estimates the average value of the intrinsic depth perception level: (μ = 5,263 ; σ = 2,42) while the standard deviance estimates the effect of the average aggregated noise affected by the individual scales of the observer persons (a).

The individual differences can be attributed to the linear distortion (parameters a and b) for every person. As the noise was supposed to have zero average (the DC component is set to b) we could estimate these values for every person and compensate their results.

The following diagrams show the point estimations of the corrected answers.

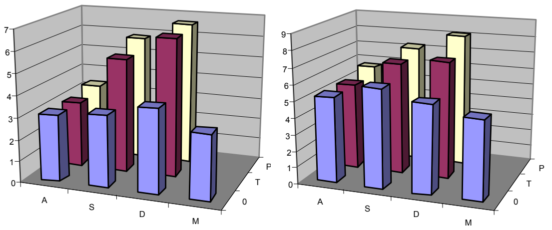

Fig. 10: Average of compensated answers for non-rotating (left) and rotating (right) scenes

Table 2: Point estimation of F

F(X1,X2,N)	**0**	**T**	**P**
A	3,08	3,04	3,30
S	3,32	5,29	5,83
D	3,87	6,44	6,64
M	2,97	-	-

σ(X1,X2,N)	**0**	**T**	**P**
A	0,66	1,54	1,49
S	1,63	1,87	2,12
D	1,75	2,72	2,13
M	1,09	-	-

F(X1,X2,R)	0	T	P
A	5,23	5,22	5,73
S	5,99	6,78	7,17
D	5,38	7,13	8,10
M	4,79	-	-

σ(X1,X2,R)	0	T	P
A	1,16	1,50	1,26
S	1,73	1,36	1,89
D	1,49	1,90	1,37
M	0,84	-	-

It was expected that we experienced the smallest value at F(M,0,N)=2,97. Surprisingly, the anaglyph technique has performed overall very weak compared to [10].

It is also interesting that the observer persons claimed better 3D quality if they wore some kind of 3D glasses but the scene were monoscopic in real. Shutter technique improves 0.4, dual stereo improves 0.9 in monoscopic show. We have to use this result very carefully: this is the value of the illusion that one has 3D glasses on.

Shutter and polarized dual techniques improve 2.0-2.7 at different eye models. The best results have dual-stereo technique; reason is that the 2x30 Hz picture shutter frequency might be disturbing.

Very interesting is that the scene rotation increases 3D depth sensation almost equally to the 3D techniques. It improves 1.5-2.7 compared to the static scenes!

Investigating the split of the results one can state that the small and big results have smaller, the middle results have greater standard deviance. The reason might be that the observer persons agreed in the obviously good and bad depth sensation results and were split by the middle ones but this effect can be interpreted with the saturation of the representation scale.

According to both variance and means values we can state that parallel camera model was more successful at projected scenes than the convergent model.

5.3.2 Evaluation of the comparison of adjacent scenes

In this case the answers lay also between 1 and 9; 5 was the neutral answer. Thus we transformed the results into the ±4 interval.

The expected values of the individual answers are not the same therefore we have to perform another type of pre-processing. All these answers are related to an edge of the adjacency graph; if we check an opposite direction we expect the opposed value. Therefore we can build a statistics from these values with corrected sign.

The results can be summarized like in the followings. The motion improves 2.05 (± 0.39 @ 90%) compared to the static scene. The binocular cue produces by 1.91 (± 0.56 @ 90%) greater values than the monocular ones. This value was 1.83 (± 0.81 @ 90%) for static scenes and 2.0 (± 0.83 @ 90%) for rotating scenes.

6 References

[1] Sheridan, T., *Telerobotics, Automation, and Human Supervisory Control*, MIT Press, Cambridge, MA, 1992

[2] L. Vajta; T. Juhász: *The Role of 3D Simulation in the Advanced Robotic Design, Test and Control* International Journal of Advanced Robotic Systems – Cutting Edge Robotics 2005, ISBN 3-86611-038-3; pp. 47-61.

[3] B. G. Blundell: *Enhanced Visualization*, Wiley-Interscience, 2007

[4] David F. McAllister. *Display Technology: Stereo & 3D Display Technologies*, Wiley Encyclopedia on Imaging, Jan., 2002, pp. 1327-1344

[5] http://en.wikipedia.org/wiki/Anaglyph_image – from Wikipedia, the free encyclopedia [07.06.2007]

[6] http://www.barco.com/entertainment/en/stereoscopic/infitec.asp

[7] S. St-Laurent: *The COMPLETE Effect and HLSL Guide*, Paradoxal Press, September 2005

[8] R. Fosner: *Real-Time Shader Programming*, Morgan Kaufmann Series, December 2002

[9] J. McVeigh, V. S. Grinberg, and M. W. Siegel. *A double buffering technique for binocular imaging in a window*, In Proceedings of the 1995 SPIE/IS+T Conference (San Jose), pages 168-175, Bellingham WA, February 1995

[10] S. Volbracht, G. Domik, K. Shahrbabaki, G Fels, *How effective are 3D display modes?* - Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 540-541, Atlanta, Georgia, US, 1997

[11] G. S. Hubona, P. N. Wheeler, G. W. Shirah, M. Brandt, *The Relative Contributions of Stereo, Lighting, and Background Scenes in Promoting 3D Depth Visualization*, ACM Transactions on Computer-Human Interaction (TOCHI), Volume 6 , Issue 3 (September 1999), pp. 214-242.

[12] S. Livatino, F. Privitera, *3D Stereo Visualization for Mobile Robot Tele-Guide*, Proceedings of Robotics and Applications RA 2006, Track - Robot Applications, pp. 240-243., 2006