Evaluation and Improvements of 3D Technology for Videoconferencing

Hadi Rizek

Supervisor, Linköping University
Dr. Karljohan Lundin Palmerius

Supervisor, Acreo Swedish ICT
Prof. Kjell Brunnström
Copyright

The publishers will keep this document online on the Internet – or its possible replacement – for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: http://www.ep.liu.se.
Abstract

This thesis presents different solutions for enhancing a prototype for a 3D video conferencing system, and investigates their advantages and disadvantages. The setup of the system consists of two cameras and an autostereoscopic display hence eliminating the need to use any eyewear. The autostereoscopic display shows seven successive stereo pairs, therefore the 3D content needs ideally to be eight images captured from different views. The system uses a simple algorithm that shifts each image horizontally with a specific value for each view.

The second part of the work is to perform a subjective testing for the system. The ITU-T standards P.1301 were used as a reference for the test. The system was compared to 2D traditional conferencing systems and the improvements presented in this work were also included in the test to evaluate the reality of depth perception in the image.
Acknowledgement

I have carried out extensive research to learn more about stereoscopic technology and subjective testing. As well, I needed a great deal of supervision and help from some experts to whom I would like to express my gratitude.

Dr. Karljohan E. Lundin Palmerius, ITN, for good advice and suggestions, and for allowing me to take a look at some of his code as guidelines for testing the Kinect device. Prof. Kjell Brunnström, Acreo Swedish ICT and Dr. Mathias Johanson, Alkit Communications, provided me with good guidance, the necessary literature and all practical details. Marjorie Carleberg, Linköping University, helped with the language check of this report. All other people in Acreo Swedish ICT who have helped me to carry on with this work, thank you all!

Norrköping March 2013

Hadi Rizek
Table of Contents

1. **Introduction** .................................................................................................................. 10
   1.1. Background .................................................................................................................. 10
   1.2. Purpose & Aim ............................................................................................................. 10
   1.3. Problem Formulation .................................................................................................. 10
   1.4. Scientific Approach .................................................................................................... 11
   1.5. Limitations .................................................................................................................. 11
   1.6. Outline ......................................................................................................................... 11

2. **Theory & Facts** ............................................................................................................. 12
   2.1. Stereoscopy .................................................................................................................. 12
   2.2. Stereoscopic Displays ................................................................................................. 12
       2.2.1. *Colour Multiplex* ............................................................................................... 12
       2.2.2. *Time Multiplex with Active Shutter Glasses* ....................................................... 12
       2.2.3. *Polarization Multiplex with Passive Glasses* ....................................................... 12
       2.2.4. *Autostereoscopy* ............................................................................................... 13
   2.3. Integral Imaging ......................................................................................................... 14
   2.4. Holography ................................................................................................................. 14
   2.5. Volumetric Displays ................................................................................................... 14
   2.6. Confere ....................................................................................................................... 14
       2.6.1. *View Synthesis* ................................................................................................. 15
       2.6.2. *Multi-view Rendering Algorithm* ..................................................................... 15

3. **Method & Implementation** ........................................................................................... 16
   3.1. Depth Camera (Microsoft Kinect) .............................................................................. 16
       3.1.1. *Holes and Unstable Boundaries* ....................................................................... 16
       3.1.2. *Solution for Multi-Kinect Interference Problem* ............................................... 17
   3.2. View Synthesis .......................................................................................................... 17
       3.2.1. *Interpolation vs. Extrapolation* ......................................................................... 17
       3.2.2. *Disparity Map* ................................................................................................. 18
3.2.3. Synthesize Views ................................................................. 18

3.3. Multi-Camera System .............................................................. 20

3.3.1. Prototype for the Multi-Camera system .................................. 20
3.3.2. Multi-Camera Setup ............................................................ 22

3.4. Subjective Evaluation .............................................................. 23

3.4.1. Test Conditions and Experiment Design ................................. 23
3.4.2. Tasks ................................................................................. 25
3.4.2.1 Ball-Drop Task ................................................................. 26
3.4.2.2 Judge-Distance Task .......................................................... 26

3.4.3. Subjects .............................................................................. 26

3.4.4. Ambient Room and Equipment Characteristics ..................... 27

3.4.5. Data Analysis ...................................................................... 27

3.4.5.1 Mean Opinion Score-based Questions ............................... 27
3.4.5.2 Depth-Based Tasks Tasks .................................................. 28

4. Results & Discussion .................................................................. 30

4.1. Depth Camera ......................................................................... 30
4.2. View Synthesis ........................................................................ 30

4.3. Multi-Camera System .............................................................. 32

4.3.1. Prototype for the Multi-Camera System ............................... 32
4.3.2. Multi-Camera Setup ............................................................ 32

4.4. Projection Problem .................................................................. 32

4.5. Viewing Distance Problem ...................................................... 33

4.5.1. Solutions for better viewing distance ................................. 33

4.6. Mismatch Problem .................................................................. 33

4.7. Data Analysis for the Subjective Evaluation .............................. 34

4.7.1. Questionnaire ..................................................................... 34
4.7.2. Ball-Drop Task .................................................................. 36
4.7.3. Judge Distance Task ............................................................ 38
# Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Different views for an autostereoscopic display</td>
<td>13</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Illustration of lenticular lenses</td>
<td>13</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Confero graphical user interface</td>
<td>14</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Microsoft Kinect</td>
<td>16</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Depth map captured from Kinect device</td>
<td>16</td>
</tr>
<tr>
<td>Figure 6a</td>
<td>Extrapolation; The two real cameras in green and the virtual camera in red which is outside. Occluded area is shaded in black.</td>
<td>18</td>
</tr>
<tr>
<td>Figure 6b</td>
<td>Interpolation; The two real cameras in green and the virtual camera in red which is in between. Occluded area is shaded in black.</td>
<td>18</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Baseline for the eight cameras. Interpolation setup, the two outermost cameras are the real cameras and the virtual cameras are placed in between.</td>
<td>19</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Rendered test scene with NVidia Mental Ray renderer</td>
<td>20</td>
</tr>
<tr>
<td>Figure 9a</td>
<td>Illustrates the stereo setup with two cameras and the comfortable area (in green) and the uncomfortable area (in red).</td>
<td>21</td>
</tr>
<tr>
<td>Figure 9b</td>
<td>Illustrates viewing area for each view.</td>
<td>21</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Parallel axis method, the left image presents the setup of the cameras, right image shows a rendered scene. Less keystone distortion appears even for the teapot which lies within the comfortable area.</td>
<td>21</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Toe-In method, the left image presents the setup of the cameras, right image shows a rendered scene. Less keystone distortion appears for the teapot which lies within the comfortable area.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Off-axis method, the left image presents the setup of the cameras; right image shows a rendered scene. Less keystone distortion appears for the teapot which lies within the comfortable area and even for the two other boxes which are outside the comfortable area.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Image to the left illustrates the Toe In method for eight cameras. Right image shows the eight cameras that have been used for the system.</td>
<td>23</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Illustration of Multi-Camera system</td>
<td>23</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Equipment for the Ball-drop task</td>
<td>26</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Equipment for the Judge-Distance Task; right image shows the vertical distance (Red-Blue) is bigger than the horizontal distance (Blue-Blue) while the distance is actually equal, left image.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 17a</td>
<td>A Histogram that presents the mean value of Q1 for each system.</td>
<td>28</td>
</tr>
<tr>
<td>Figure 17b</td>
<td>Error bars for mean value of Q1 for each system.</td>
<td>28</td>
</tr>
<tr>
<td>Figure 18</td>
<td>Results captured from the Kinect device, the right image presents the depth image with depth value per pixel. The left image presents the depth image textured with RGB-values captured by the colour camera.</td>
<td>30</td>
</tr>
<tr>
<td>Figure 19</td>
<td>The left image of a stereo image pair, and its corresponding disparity map calculated using MATLAB disparity function.</td>
<td>31</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Timeline of the view synthesize progress. DIBR blocks represent the process of Depth Image Based Rendering. IL’ and IR’ are the calculated candidate images from the original image and its corresponding disparity map.</td>
<td>31</td>
</tr>
<tr>
<td>Figure 21a</td>
<td>illustrates the projection problem that occurs by using the Toe-in method.</td>
<td>33</td>
</tr>
</tbody>
</table>
**Figure 21b** illustrates the case when using the Off-Axis method…………………………………….. 33
**Figure 22** A screen shot from the display to illustrate the viewing distance…………………………………….. 33
**Figure 23** Left image shows the camera setup with larger mismatch when the convergence plane is closer to the camera’s baseline, while the image on the right shows a smaller mismatch when the convergence plane is further away from the camera’s baseline……………………………………….. 34
**Figure 24** The mean value of ball score compared between the two systems for each test person, where 2D presents the monoscopic system and 3D presents the autostereoscopic system…… 37
**Figure 25** The mean value of guesses for the Judge-Distance Task for the two systems…………………..38
**Figure 26** The amount of different answers in percentage, the total amount is 26 answers……………39
**Figure 27** The amount of answers for which system was more preferable, the total number of answers is 26……………………………………………………………………………………………………… 39

**Tables**

**Table 1** Time table for subjective testing………………………………………………………………………. 24
**Table 2** Setup for the test of the monoscopic system…………………………………………………………… 24
**Table 3** Setup for the test of the autostereoscopic system ……………………………………………………… 25
**Table 4** Mean Opinion Score ……………………………………………………………………………………... 27
**Table 5** Comparison for the mean value for each question between the two systems………………….. 34
**Table 6** Results from the Ball-Drop Task……………………………………………………………………….. 36
1. Introduction

Today’s traditional 2D video conferencing systems cover most of the cues used by human to visualize the world’s 3D structure, such as occlusion, perspective and familiar size. That is why users can make sense of images and video. What are missing are the following four eye mechanisms

- **Binocular disparity**, seeing a different image with each eye.
- **Movement parallax**, seeing different images when changing viewing point.
- **Accommodation**, the eyes’ lenses focus on the object of interest.
- **Convergence**, both eyes converge on the object of interest.

These cues deliver to the users the sense of presence and give them the feeling of being in the same space while being geographically distributed. Therefore, traditional 2D videoconferencing does not replace physical conferencing.

1.1. Background

Alkit Communication AB is a Swedish company which develops communication systems and solutions. In a research project together with Acreo Swedish ICT AB, a Swedish research institute, a prototype system for stereoscopic video communication, Confero Telepresence HD3D has been developed, whereby 3D videoconferences can be realized. Acreo is launching a study of stereoscopic videoconferencing based on Alkit’s system. The current setup of the system consists of two Full-HD cameras and an autostereoscopic display thus there is no need for any eyewear to achieve stereopsis. The display presents seven stereoscopic views and needs therefore input from eight images taken from different viewpoints. The two cameras provide the system with only two video signals which are shifted horizontally for each view. The result of this is poor depth perception and possibly increased visual discomfort [1] for the views far from the centre image. A subjective test is also needed to evaluate the system and compare it to traditional videoconferencing.

1.2. Purpose & Aim

The purpose of this work is to present different solutions to enhance the system with an efficient method that can present eight different views. As a result, better depth perception can be provided to the user. Many aspects need to be considered while constructing the method: the system is limited to the bandwidth provided; the resolution of each image will be divided by eight and the viewing distance of the display. The aim of the work is to arrive at a solution where 3D videoconferencing is perceptually preferable to traditional 2D videoconferencing.

1.3. Problem Formulation

The method that is used for the current system to synthesize the different views delivers poor depth quality and non-realistic views. There are two main questions here that need to be answered.

- Is there any other method that can deliver better depth quality and more realistic views?
- Does stereoscopy bring any added value to videoconferencing?
1.4. Scientific Approach

To answer the questions in 1.3. the work was divided in two main parts.

- The first part consists of research and investigation of several methods for generating different views that delivers better quality of depth perception and more realistic views.
- The second part is performing user study for the system and comparing it to traditional videoconferencing.

1.5. Limitations

The report investigates different solutions for generating the different views for Confero. There are many limitations for the system in hardware and software.

- **Hardware:** The recommended viewing distance for the autostereoscopic display is four metres which is generally a large distance for videoconferencing. The cameras that have been used for capturing several views are not advanced and deliver non-identical image quality.
- **Software:** There is a delay in the video signal for the autostereoscopic system due to un-optimized rendering algorithm, and that affects the sense of presence.

Due to these limitations the results of this work may therefore not be generalized for all autostereoscopic videoconferencing systems.

1.6. Outline

The work has started with research and investigation of the different technologies related to autostereoscopic videoconferencing, some of the keywords were: stereoscopy, autostereoscopy, 3D displays, videoconferencing and subjective evaluation. After gathering a huge amount of knowledge in this area the work has started with looking into the system itself and trying to understand it and see the way it works. All the main technical background information, theory and facts are presented in Theory & Facts (2).

The second part of the work was trying to find more efficient solution for presenting 3D using the autostereoscopic display, the different methods that have been investigated are: depth camera (3.1), view synthesis (3.2) and multi-camera system (3.3). The results of this investigation are presented in Results & Discussion (4.1, 4.2, 4.3).

The last part was to perform a subjective evaluation for the system (3.4) and compare it to traditional videoconferencing, several test persons (subjects) were invited to do the test. They have performed few tasks and answered a questionnaire regarding the different aspects that were evaluated; the answers were based on their experience of each system. The results of the subjective evaluation have been analysed and presented in 4.7.
2. Theory & Facts

This chapter contains brief explanations for the different technologies that are related to this work, it also includes a description for the videoconference system, Confero.

2.1. Stereoscopy

The principle of stereoscopy is to present an image to one of the eyes and another horizontally shifted version of the same image to the other eye. The human brain merges these two images into a single view and creates a sense of depth. In a stereoscopic presentation each eye will only see the corresponding image. This is traditionally carried out by using different kinds of special eyewear (aka 3D glasses, eye-glasses or anaglyphs). Different displays cause different levels of visual fatigue and visual discomfort which is clarified more by e.g. Lambooij and IJsselsteijn [1]. The different types of stereoscopic displays are described in 2.2

2.2. Stereoscopic Displays

There are different technologies for presenting stereoscopy; the most common ones are explained here.

2.2.1. Colour Multiplex

Colour multiplex display is based on putting each of the stereo image pair in a certain colour channel, for example red and cyan. By the use of colour filters, each of the images will be seen only by one eye and the brain composites the two images and creates the sense of depth in the viewed image. Anaglyph glasses are cheap and the method for generating anaglyph stereo is simple and works for any 2D colour display. Unfortunately, the resultant image has poor quality and poor colours so, as a result, the use of this technique is very rarely used in today’s 3D video production.

2.2.2. Time Multiplex with Active Shutter Glasses

Using the time multiplex method, the two images are received by the human eyes at different points in time. The display switches between the left and right image with a small time difference that cannot be perceived. Active shutter glasses are used to block the view for one of the eyes when the other eye is seeing its corresponding image. One condition, the updated frequency needs to be higher enough; the frequency per eye should be higher than flicker-fusion-threshold which is equal to 50-60 Hz. Flickering and low luminance are problems that occur when using active glasses. Another problem is crosstalk that may occur due to different timing issues between the shutter glasses and/or a slow response time of the LCD display. Active shutter glasses are quite expensive, bulky and need batteries.

2.2.3. Polarization Multiplex with Passive Glasses

As opposed to the time multiplex principle, the images at the polarization multiplex principle are displayed at the same time. The passive glasses contain a polarization filter that ensures that each view is only seen by the corresponding eye. Passive glasses can contain vertical and horizontal polarization filters or clockwise and anti-clockwise circular polarization ones. Since the two images are interleaved
on the display at the same time, half of the resolution is only presented for each eye and so that decreases the quality of the perceived video [3].

2.2.4. Autostereoscopy

An autostereoscopic display provides depth perception without requiring any stereo glasses. There are three different types of autostereoscopic displays.

- Passive two-view display (No tracking)
- Active two-view display (head-tracking)
- Passive multi-view display.

For the two-view display only two views are provided and the user is limited to a fixed view point called a “sweet spot” for watching the video scene, see Figure 1. The two different images are cut into strips and interlaced together, and an array of lenticular lenses or parallax barriers is placed in front of the display to separate the different views, Figure 2. Within the sweet spot each eye receives its corresponding image and the illusion of depth will be created in the brain. Active autostereoscopic displays are more efficient in term of adjusting the position of the sweet spot to the position of the viewer’s head and reducing the chance for the user to end up in the wrong view area. This adjustment is carried out by mechanically shifting the lens array or the parallax barrier in front of the display or using a liquid crystal display to form the parallax barriers. However, the problem here is not shifting the view area but the tracking system itself, since it would be senseless to replace stereo glasses with a head mounted tracking device, and the system would be limited to only one user.

Multi-view displays present a more realistic image where the user can see different views when he angles or turns his head. Under the lenticular lens in the multi-view display there are a specific number of pixels coming from the same number of different single views, i.e. eight views gives eight pixels. Theses pixels are diverged by the lenticular lens and that enlarge the spatial difference in geometry which results in different views presented for the viewer. Using this display makes it possible for multiple viewers to perceive a stereo image when they are standing anywhere within the viewing area. It also provides different viewpoints, which means the ability to look around an object in a limited angle by the movement of the viewer’s head.

There are different technologies for constructing a multi-view autostereoscopic display.

- **Time sequential** where all views are displayed on a single very fast display.
- **Multi-projector** where for each view a single projection display is used.
- **Spatial multiplex** where the different views are displayed by splitting the resolution of the display.

The display used in this project is based on spatial multiplex technology. A sheet of lenticular lenses is placed in front an LCD display so the resolution is divided between eight views. The lenticular lenses refract...
the light coming from the underlying sub-pixels causing dark zones between viewing slots. By diverging these pixels, the lenses enlarge the spatial differences in geometry.

2.3. Integral Imaging

A true three-dimensional image is achieved by integral imaging, also called integral photography. It has a capture stage where the image of an object is recorded in a pickup plate. The lenticular lenses produce elemental images in front of each lens in the same way as the viewing zones of autostereoscopic display in 2.2.4 are gained. Each elemental image retains \( P_e \) pixels, the total pixel count \( P_t \) is calculated is in equation (1)

\[
P_t = P_e P_l
\]

Where \( P_l \) is the number of lenses. The image quality is improved by multiplying the elemental images pixel count by the number of pixels which is usually larger than six for an extended side view of the object. The result is a large number of pixels compared to a traditional two-dimensional display, and that is the challenge for integral imaging.

2.4. Holography

Based on the physical properties of light, holography can provide a true full parallax three dimensional display. The pixel of the holographic display emits light beams in different directions with different intensities and colours. A surface that emits light can be composed of these pixels and will act as a hologram and be able to show three dimensional images.

2.5. Volumetric Displays

Different from stereoscopic and autostereoscopic displays that shows two images, one for the left eye and one for the right eye, volumetric displays provide one image for both eyes. The 3D effect is achieved by applying an array of images on several planar displays placed on top of each other; each is at a different distance or depth from the eye. Volumetric display refers to the volume of these multi-depth displays. There are two different technologies; static volume display where the stack of displays is stationary, and swept volume display where the third dimension is achieved by using mechanical displays, which have a rotational or translational movement through the volume while images are emitted at different depths.

2.6. Confero

Confero is a videoconference application that supports real-time multimedia communication between two or more users. What makes Confero different from other monoscopic videoconferencing applications is its ability to present three-dimensional depth perception. That is done by using two vertically-aligned cameras with a horizontal separation of 65 mm which is the average distance for human eye separation. Since using stereo glasses will reduce the sense of presence, an autostereoscopic algorithm was therefore implemented and an autostereoscopic display was used for this purpose. The display

\[\text{Figure 3 Confero graphical user interface.}\]
supports eight views and thus needs an input of eight images with different viewpoints. Since the system is based only on two cameras, the rendering algorithm proposed by Johanson [4] has been applied to synthesize the different views.

2.6.1. View Synthesis

The view synthesis algorithm is based on mapping the two original video streams as the centremost channels of the screen. The six other views are generated simply by horizontal translation of the input stream, this translation being consistent with the estimated head-motion of the user. The weakness of this algorithm is the estimation of the user’s head-motion which cannot be exact and hard to calculate. The generated views contain the same information as the general input stream and no more information is presented as it would be in reality when looking at a scene from different viewpoints.

2.6.2. Multi-view Rendering Algorithm

Equation (2) presented in (Berkel 1999) [5] is used for multi-view rendering. \( N \) is used to assign the appropriate image data to each pixel.

\[
N = \frac{(k+k_{\text{offset}}-3l \tan \alpha) \text{mod} x}{x} N_{\text{tot}}
\]  

(2)

\( k, l \) point to individual red, green or blue subpixels, and are calculated from the pitch of each pixel \( P_h \) and the pixel coordinates \( x, y \) according to equation (3a)

\[
k = \frac{x}{P_h}, \quad l = \frac{y}{3P_h}
\]  

(3a)

\( N_{\text{tot}} \) is the total number of views and \( \alpha \) is the rotation angle of the lenticular lens. The parameter \( k_{\text{offset}} \) accommodates an arbitrary horizontal shift of the lenticular lens array with respect to the LCD. The parameter \( X \) presents the number of views per lens. It is calculated according to equation (3b),

\[
X = \frac{m+1}{m} \cdot \frac{P_{\mu}}{P_h \cos \alpha}
\]  

(3b)

\( P_{\mu} \) is the lens pitch and \( m \) is the magnification that can be expressed by the viewing distance \( D \) and the focal depth \( f \) of the camera, \( m + 1 = f D \).
3. Method & Implementation

This chapter presents three different techniques for generating different camera views, these methods have been investigated in order to find an efficient method that delivers a result which can be used for the autostereoscopic videoconference system. The first method is uses a depth camera (e.g. Microsoft Kinect) to get depth information from the scene. Afterwards the scene can be reconstructed as a three-dimensional geometry. Another method is to synthesize the different views using fewer cameras, at least two or more. Finally, using real multiple cameras gives results that can be used as reference for further development of the system in the future.

3.1. Depth Camera (Microsoft Kinect)

A depth camera is a camera that provides a depth image with a depth value for each pixel. These values can be used to reconstruct the scene in a three dimensional coordinate system. Microsoft Kinect is relatively cheap version of a depth camera and can be used for this purpose. Kinect is built on a system that can interpret specific gestures, which gives a completely hands-free control of electronic devices by using an infrared projector and camera and a special microchip to track the movement of objects and individuals in three dimensions. As seen in Figure 4, the sensor is a horizontal bar connected to a small base with motorized pivot. The bar contains an RGB camera, depth sensor and multi-array microphone. The depth sensor is based on infrared technology, thus it contains an infrared laser projector combined with a monochrome IR sensitive CMOS sensor. This is available to capture any video data in three dimensions under any ambient light conditions except sunlight due to its infrared radiation which causes distortion for the Kinect’s IR pattern. The Kinect sensor provides colour image, infrared image, depth map and audio capture. The depth image provided by Kinect can be used for to reconstruct three dimensional images. Since the Kinect’s RGB camera provides poor resolution, a HD camera can be used instead of the RGB camera for a better image resolution. Care is needed to obtain good calibration between the IR camera and the HD camera. Using the IR camera, a 3D geometrical mesh can be reconstructed and textured using the image captured by the HD camera. The mesh can afterwards be rendered onto the display with the correct right/left frustum installed depending on the viewer position. The results will later be assembled into the pattern required by the autostereoscopic display.

3.1.1. Holes and Unstable Boundaries

The black areas in Figure 5 defined as holes are either occluded by other objects or absorb IR light. Additional noise occurs when trying to observe the projected light on a surface which is perpendicular to the IR camera. The results are unstable boundaries of objects and the poor alignment of depth and colour edges. Therefore an in-painting technique and noise filtering need to be applied before the depth image is useful for further visual work. However, several facts and
phenomena need to be considered regarding the Kinect camera:

- Kinect depth range is limited to 0.5-3.5 m
- Kinect depth error can be reduced if the camera is near the line of sight of the user.
- The IR camera and HD camera need to be calibrated.
- Holes are black areas which are undefined by the IR camera and therefore have no depth information.
- Unstable boundaries caused by observing the light reflected from perpendicular surfaces to the IR Camera.

Many of these problems can be solved using several different cameras [6]. The problem is that the Kinect devices will project the same dot pattern at the same wavelength, thus it will be hard for each of them to distinguish its own dot pattern. This interference problem causes holes and small amount of high frequency depth noise. Several solutions are proposed in [6], [7] and explained in 3.1.2

3.1.2. Solution for Multi-Kinect Interference Problem

When interference occurs between two or more Kinect devices, the unit generally does not return any data. A lack of depth data can also occur due to the offset between the IR projector and the IR camera.

1. **Software solution:** Assuming that small holes are a part of the continuous surface, these holes can be filled by colour interpolation or another method. However, large holes are ignored and hopefully are captured by the other unit.

2. **Hardware solution:** The idea here is to make sure that the device only sees its own projected dot pattern. This can be achieved by applying synchronized shutters on each unit’s IR projector and camera, but this will unfortunately reduce the frame rate and amount of light seen by the IR camera. Another more efficient approach is to apply a small amount of motion on the device, and that will make the device’s own dot pattern appear sharper than the other’s dot pattern. By using high special frequency patterns, each sensor will see each pattern with a higher contrast. The colour image will also be blurred for the moving units, therefore it needs to be sharpened in the post-processing phase.

3.2. View Synthesis

View synthesis is a way for generating different views based on two source images taken from two different viewpoints.

3.2.1. Interpolation vs. Extrapolation

The most common method for view synthesis is to interpolate the views between two outermost cameras. Figure 6a shows the two cameras in green and the virtual camera in red. It is obvious that the extrapolated view will include areas of the scene marked in black which are not seen by the original cameras. These areas will appear as holes in the synthesized image. View interpolation will therefore result in synthesized views with a fewer number of holes as seen in Figure 6b.
3.2.2. Disparity Map

To be able to synthesize the different views, the disparity map for both images needs to be calculated. There are many methods available to calculate the disparity map, and there are several research publications regarding this problem (e.g. Mühlmann et. al. 2002 [8]). A disparity map includes information about the pixel difference or motion between a pair of stereo images; objects that are closer to the camera baseline have a bigger horizontal shift than objects that are further away. The core problem here is to find corresponding points in the two source images; the better one can do the better the disparity map will become. Mismatching between pixels in the two views causes holes and disconnected edges in the resultant depth map.

3.2.3. Synthesize Views

Assuming that the disparity map for the stereo image pair is given then the synthesized views can be carried out in different methods. A common and effective method is Depth Based Image Rendering (DBIR), as in equation (4) presented in [9]

\[ Z' \hat{n}' = ZA'R^{-1}\hat{n} + A't \]  

(4)

Where \( \hat{n}' \) and \( \hat{n} \) represents the two 2D image points in homogeneous notation. The equation contains intrinsic and extrinsic properties for the cameras. The intrinsic matrix \( A \) contains the focal depths in pixels since it is given in distance unit (e.g. metre) originally. The image coordinates could be translated from the image centre that is adjusted by \( p_x \) and \( p_y \). The perpendicularity of the axis in the image can be modelled as a skew parameter \( s \), that is used to skew the frustum of the camera,
The extrinsic matrix $K$ consists of a rotation matrix $R$ and translation vector $t$ of the camera relative to the world surrounding. An easier way of thinking is to imagine the opposite, the camera is fixed and the world rotating and translating in front of it.

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \quad t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

(6)

$$K = [R|T] = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{pmatrix}$$

(7)

This is carried out by applying a 4x4 matrix transformation to the point in the world equation (8)

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f m_x & 0 & s & p_x \\ 0 & f m_y & 0 & p_y \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

(8)

$w$ is the homogenous coordinate of the image plane. Figure 7 shows the setup of the cameras, the outermost cameras, coloured in green, illustrate the real cameras, and the red ones in between are the virtual cameras. The separation distance between each of these two cameras is equal to the average of the intraocular eye distance about 65 mm [4], and each of them works as a right and left camera at the same time depending on the position of the viewer’s head.

Figure 7 Baseline for the eight cameras. Interpolation setup, the two outermost cameras are the real cameras and the virtual cameras are calculated in between.
Using the input from the two real cameras with corresponding disparity maps two candidate images can be calculated (left candidate and right candidate) for each view. The candidate images can be merged under the following conditions.

- If a pixel is seen by both cameras, the pixel with higher disparity value is chosen
- If a pixel is seen by one camera only, this pixel is chosen
- If a pixel is not seen by both cameras, it is flagged as a HOLE and given a black or white colour.

The next step is filling the holes. There are different methods that can be applied. Interpolation of the background colour is a possibility; however, it might result in smoothed areas. Another way is to apply a Gaussian filter which is basically a low-pass filtering, that will cause distortion in the geometry of the image. Inpainting of colour that is based on information from the depth map is another solution; the disadvantage here is generated artefacts in the filled areas.

Xi. Ming, et al. [7] presents the ”Oriented examplar-based inpainting” method. They define Hole-area and Visible-area in the scene. For each pixel in the Hole-area a priority value is assigned depending on how close the pixel is to the edge of the hole (the closer the pixel is, the bigger priority value is has). The pixels are filled afterward with the best match patch in the Visible-area.

### 3.3. Multi-Camera System

Using a multi-camera system requires research for the setup of the cameras which considers the requirements for the autostereoscopic display such as the viewing distance and number of views. There are also different techniques for a stereoscopic camera setup: usually the setup is applied for two cameras but in this case there are several cameras involved, and therefore a prototype for the multi-view system needs to be produced to see the different results of each method and work out which multi-camera setup is the most efficient for this purpose.

#### 3.3.1. Prototype for the Multi-Camera system

The prototype was created by using 3D studio Max 2012¹, and presents the three different techniques used for generating stereoscopic video sequences. In the three different techniques the separation between the cameras is equal to the average eye separation distance which is 65 mm and that makes the eight cameras baseline distance 455 mm. That means for each view the spectator will have a view angle of 6.51 degrees, where the view distance is 400 cm. The scene was constructed with three different objects, a teapot and two boxes. The teapot was placed in the center of the convergence plane, which means in the

---

¹ http://usa.autodesk.com/3ds-max/
middle of the comfortable area, see Figure 9a. The other two boxes were placed behind the teapot, where the green box borders the comfortable area and the blue box is in the uncomfortable area.

In the Parallel-Axis method the cameras are offset from the center axis with eye separation distance kept constant between each two cameras. The viewing axis for the cameras is parallel and the projection planes are in a parallel projection alignment. Figure 10 presents the camera setup and the keystone distortion that occurs using this method. Notice that the figure shows a 2-camera system.

In the Toe-In method the cameras are offset from the center axis and angled inward towards the intersection point between the center axis and the convergence plane. The projection planes do not align and that makes it difficult to converge at the edges of the screen where keystone distortion appears.
The **Off-Axis** method is similar to the Toe-In method; the difference here is that the projection planes are aligned by skewing the frustum of each camera. That gives better results. An 8-Camera plug-in provided by Alioscopy\(^2\) is used here to generate the system; the plug-in is compatible with the autostereoscopic display.

---

**3.3.2. Multi-Camera Setup**

The setup of the multi-camera system is eight HD cameras with a resolution of 1280x720 pixels. Calibrating eight cameras is not an easy task and an additional one is to toe-in the cameras into the centre of the convergence plane as shown in figure 13

---

\(^2\) [http://www.alioscopy.se](http://www.alioscopy.se)
The calibration of the cameras has been carried out manually and a construction has been built to keep the cameras in the same horizontal line with the ability to rotate each camera separately around a vertical axis through the lens and, by that, being able to toe-in the cameras into a given point, see figure 13.

3.4. Subjective Evaluation

A subjective test has been carried out for the autostereoscopic videoconference system; the test covered the experience of audio-visual quality and depth-perception. It included comparing the traditional 2D tele-meeting with the multi-camera system. The result has been analysed and a conclusion was reached about how further improvements could be implemented.

3.4.1. Test Conditions and Experiment Design

The test was divided into two different sessions. Since the test is time consuming the total time needed to be reasonably compressed to minimize participant fatigue and maximize the accuracy of the experiment. However, each conversation was at least five minutes long [10]. Each test was carried out for a pair of subjects at the same time according to Table 1.
In the first session, traditional 2D-conferencing was presented to the subjects to compare it with the other methods involved in the test. The setup of the test was 2D-LCD display and a HD camera, see Table 2.

**Table 2 Setup for the test of the monoscopic system**

<table>
<thead>
<tr>
<th>Subject 1 (Room 1)</th>
<th>Subject 2 (Room 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A HD-camera and an LCD Display</td>
<td>A HD-camera and LCD monitor</td>
</tr>
</tbody>
</table>

In the second session, the autostereoscopy with 8-cameras setup was presented, see Table 3. The ideal case is to present an image with full-HD resolution, but due to the limitations mentioned in 1.5. The video quality for the autostereoscopic system was low compare to the video presented on the LCD display.
The two different technologies were compared to evaluate the quality of depth perception in videoconferencing and to see if it adds any value for immersive presence. Since the viewing distance of the available autostereoscopic display was four metres, the same distance was used for the other mode.

3.4.2. Tasks

The tasks have been designed according to ITU Recommendations P.805 [11], the following guidelines are provided for the design of task-based tests:

- The tasks need to be designed so the subjects maintain their attention on the audio-visual terminal during the conversation.
- The tasks need to be designed so they can be performed by a wide range of subjects including elderly and hearing-impaired subjects.
- Each task need to be preferably performed by a pair of subjects to achieve a real-life audio-visual communication. However, the test leader is not advised to participate in the test.

Several tasks were used for the subjective test to evaluate the system and compare it to monoscopic videoconferencing. The tasks that were provided by ITU-Documentation were focused on 2D telecommunication and several aspects such as delay and package loss which do not concern the purpose of this test. Therefore, new tasks which are depth-based needed to be designed to evaluate the two features provided by the autostereoscopic system, depth perception and multi-view.
3.4.2.1. **Ball-Drop Task**

This task is based on collaboration between the test subject and the other party. Let us call the test subject Person A and the other party assisting Person B. Person B is provided with three balls and a table marked with horizontal lines dividing the table in eight different sections (white lines), three other vertical lines divided the table in three different fields (black lines), see Figure 15. In each field a marker was placed in a different part from the other field. Person B is asked to hold a ball randomly over these fields, and for each time the Person B holds the ball over the table, Person A needs to judge if the ball will hit the marker if Person B drop it. If Person A says yes, Person B will simply let the ball fall down and see if it hits the marker. If Person A says no, then Person B will take the ball back and place it somewhere else over the table. The sections were weighted with different values with the highest value for the section with the marker and less values the further from the marker the section is. The task of Person A is to try to hit the three markers with the three balls and by these means gather as many points as possible. For each try, the point collected is registered in a specific protocol for this task. The task was performed for the two systems to compare the quality of depth perception.

3.4.2.2. **Judge-Distance Task**

A Whiteboard was placed with a small angle to the viewing axes of the camera, see figure 16. On the Whiteboard there were three magnets forming a L-shape. The middle magnet is static; the other two magnets can be moved further or closer to the middle one by Person B. Person B covers the whiteboard by standing in front of it while moving the magnets. The task for Person A is to judge the distance between the two outer magnets related to the middle one. This task was performed three times and the correct guesses were registered in the protocol of the task for further analysis.

3.4.3. **Subjects**

The number of subjects should not be less than 16 according to the ITU Recommendations. However, the exact number depends on the accuracy required for the results and will be later stated by the experimental design. It is recommended that subjects should be non-expert, and not directly involved with videoconferencing as a part of their normal work.
3.4.4. Ambient Room and Equipment Characteristics

A list of parameters is provided in [12] for typical viewing and listening conditions that are used in audio-visual quality assessment. Appendix II includes specifications for the equipment characteristics.

3.4.5. Data Analysis

The test results were divided into different parts. The first part is based on the Mean Opinion Score (MOS) test method that has been used for telecommunication subjective evaluation. The test subjects were asked to answer several questions according to the scale in Table 4

<table>
<thead>
<tr>
<th>MOS</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad</td>
<td>1</td>
</tr>
<tr>
<td>Poor</td>
<td>2</td>
</tr>
<tr>
<td>Fair</td>
<td>3</td>
</tr>
<tr>
<td>Good</td>
<td>4</td>
</tr>
<tr>
<td>Excellent</td>
<td>5</td>
</tr>
</tbody>
</table>

The second part is the result from the depth-based tasks, and the third part consists of several answers and comments the test subjects have given during the test.

3.4.5.1. Mean Opinion Score-based Questions

The questions that were asked for the participants can be found in Appendix III. Five of these questions were used in this analysis,

- Q1: How would you rate the overall audio-visual quality?
- Q2: How would you rate the video quality of the connection?
- Q3: How would you rate the quality of depth perception?
- Q4: How would you rate the reality of the virtual representation of the other person?
- Q5: In what grade did you experience that the other party was present in the same room?

The subjects were asked to answer these questions using the Mean Opinion Score as in Table 4. The same questions were asked for the two systems but in a different order. The purpose of the other questions that were included in the questionnaire was to get the general impression for the subjects. The mean opinion score for each question was calculated according to equation (9)

\[ MOS = x = \frac{1}{n} \sum_{i=1}^{n} a_i \]  

and compared between the two systems which are defined as (2D) for monoscopic and (3D) for autostereoscopic, as in Figure 17a.
The standard deviation was also calculated as in equation (10) to find out the variance of scores between the different test subjects for each question,

$$STDEV = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})$$  \hspace{1cm} (10)

This value was used afterwards to calculate the confidence interval with 95% confidence for each mean score. The limits for the confidence interval were calculated according to formula for expectation CI of a normal distribution, equation (11)

$$CI = \bar{x} \mp k.STDEV$$  \hspace{1cm} (11)

The coefficient $k$ can be calculated as $k = 1.9600/\sqrt{N}$ only for large number of samples. For small number of samples a table for coefficient $k$ [13] was used to get the value of the coefficient based on number of samples. The confidence interval was presented as error bars as in Figure 17b
The standard bars are defined by the standard error that is calculated using equation (12)

\[
\text{Std. Err} = \frac{\text{STDEV}}{\sqrt{N}}
\]  

(12)

To check the significance difference for the MOS results, a statistical T-Test was applied to compare the two mean values for each of the questions, see equation (13)

\[
T = \frac{X_1 + X_2}{\sqrt{\frac{1}{s_1^2} + \frac{1}{s_2^2}}}
\]  

(13)

Where \(X_1, s_1, X_2, s_2\) are the MOS value and the standard deviation for the monoscopic system respective the autostereoscopic system. By using the \(T\) value the probability \(p\) is calculated from Student’s T-Test with two-tailed distribution where the critical value is \(p_c = 0.05\). By comparing the \(p\) value to the critical value, one can determine whether the MOS value for the two systems is significantly different or not according to the three cases below:

- \(p > p_c\): The results are not significantly different.
- \(p < p_c\): The results are significantly different.
- \(p = p_c\): Doubtful case and no conclusion can be taken.

3.4.5.2. Depth-Based Tasks

The same procedure mentioned in 3.4.5.1. was performed to analyse the results from the Ball-Drop, and the Judge-Distance task. An additional step was applied here since each subject performed the same task: nine tries for the Ball-Drop (three balls for each field) and three times for the Judge-Distance. The average value for all nine tries for each person was calculated and the average value for all averages was used for the rest of the analyses, equation (14)

\[
\text{MOS} = X = \frac{1}{N} \sum_{i=1}^{N} x_i
\]  

(14)

Where \(x\) is the average value for \(i\)th person. The result from remaining questions and comments were compiled and presented in 4.6.4.

During the Ball-drop task each subject had three tries for each field, which means nine tries for each system. The order of tries has also been registered to see if there is any learning effect for the test subjects. Thus, there were three variances that needed to be analysed.

- The **Overall Average**, the average for all tries for each system, \(N = \) Number of subjects
- The **Field Average**, the average for each field for each systems, \(N = \) Number of subjects
- The **Try Average**, the average for each try for each system, \(N = 3\ast\text{Number of subjects}\)
4. Results & Discussion

This chapter contains the results for each method that has been implemented in this work; a discussion of the presented results is also included here. These results present answers for the questions formulated in 1.3.

4.1. Depth Camera

Using the OpenKinect\(^3\) platform, the depth data registered by the Kinect device can be mapped into a three dimensional world and rendered into the display using OpenGL. Figure 18 presents two images captured from the Kinect unit. The left image shows the colour image mapped on the depth data in a three dimensional world.

![Figure 18](image)

Figure 18 Results captured from the Kinect device, the right image presents the depth image with depth value per pixel. The left image presents the depth image textured with RGB-values captured by the colour camera.

Holes can be seen on the neck and under the nose; these areas are undefined and have no depth information. Areas that are perpendicular to the light projector such as the edge of the head are unstable and have flickering problems. Placing the hand or other object in front of the body will block the infrared rays and holes will occur again in the occluded area. Filling holes is possible in different ways but the results will include artefacts and/or distortion in the geometry of the image.

4.2. View Synthesis

The disparity map can be calculated in different ways. These images\(^4\) in Figure 19 presents two of these methods, the black areas are holes, i.e. areas which are not seen by the two real cameras. Disconnected edges can be marked by comparing the middle image and the right image to the left one. The artefacts at depth map will affect the final synthesized view strongly. The quality of the synthesized image depends also on quality of the original input video signal.

\(^3\) http://openkinect.org/wiki/Main_Page

\(^4\) The original stereo pair included is from D. Scharstein and R. Szeliski. [15]
The process of view synthesis consists of different steps, which are illustrated in Figure 20. The Depth Based Image Rendering function (DBIR) contains a translation vector which is identical for each view, thus the part of the process with grey background needs to be repeated for each view. The question that remains is using view synthesis or not?

There are many aspects that need to be considered when synthesizing views. Artefacts and distortion in the synthesized view will affect the sense of presence negatively. Calculating a satisfactory disparity map is complicated and a great deal of work needs to be carried out there. A low-quality depth map will result a low-quality synthesized view. An alternative is to use the Kinect camera to calculate the disparity map but then artefacts and holes explained in 3.1.1. will occur. As seen in Figure 20 the process is quite long and time-consuming which could work for a pre-recorded video sequence or stills. When it comes to a real-time video signal the progress will require an expensive and complicated hardware setup. Hence, all views need to be synchronized together and with the transferred audio.
4.3. Multi-Camera System

The result of this system is divided into two parts, the first part is the result from the prototype for the multi-camera system, and the second part is the result of the real multi-camera setup.

4.3.1. Prototype for the Multi-Camera System

The rendered sequences by the prototype for the multi-camera system were displayed on the autostereoscopic display using Alioscopy software which is compatible with the display. The parallel axis method gave inefficient results which might cause visual discomfort. The other two methods gave similar results to each other, where the toe-in method gave more cross talk and ghosting closer to the edge of the display. Assuming that the user will be staying close to the centre of the convergence plane, the toe-in method is a better option. Otherwise the Off-axis method is the best option but the cost will be re-computing the images for skewing the frustum of the camera.

4.3.2. Multi-Camera Setup

After calibrating the cameras, Confero was modified to define them and render the input signals using the same algorithm defined in 2.6.2. Several hardware and software problems appeared when trying to run the system, but many of them were solved either by modifying the algorithm or getting better devices, Another problem was a wrong viewing rate due to the large viewing distance which requires new cameras with zoom-in functionality and/or a new autostereoscopic display with shorter viewing distance, but, due to lack of resources, this problem was not solved. Another problem was the delay in the video signal due to un-optimized rendering algorithm, so to decrease the delay as much as possible a lower resolution has been chosen for rendering the final image, and a synthetic delay was applied for the audio signal. However, despite the delay and low resolution, the final image was acceptable and could be used later for the subjective evaluation. Unfortunately, it is impossible to present the final image in this report.

4.4. Projection Problem

Projection problems might appear while using the toe-in method. When an object such as a sphere from the outermost camera is looked at, the projected image of it should be an ellipse as shown in Figure 21b. Figure 21a illustrates the case when using the toe-in method and the projected images of a sphere is a sphere. Thus the toe-in method will cause distortion in the geometry of the image and that will be more noticeable for the outermost cameras. The bigger Toe-in angle the bigger distortion will occur.
4.5. **Viewing Distance Problem**

The recommended viewing distance for the autostereoscopic display available is four metres, which means that the distance to the convergence plane is supposed to be the same. Using cameras without a zoom-in mechanism will cause unrealistic images with the foreground much bigger than the background and the intraocular will appear very small in the background which leads to poor sense of presence, see Figure 22.

4.5.1. **Solutions for better viewing distance**

There are several autostereoscopic displays with a shorter viewing distance and a full viewing rate, so using one of these could be an option (e.g. Alioscopy 3D HD 42 CloseView). Another solution is to use zoom lenses with wide angle and telephoto setting to decrease geometry distortion and avoid keystone effect. A telephoto lens will give narrower depth of focus, but that can be fixed by using the hyper focus setting of the lens, on the other hand a shorter depth of focus can be good for bringing out the most important object in the scene. The telephoto lens has the ability to make the background appear to be closer compared to the foreground so the image will look more realistic. However, in order to present a better stereoscopy a normal focal length would be the best.

4.6. **Mismatch Problem.**

Decreasing the distance between the camera’s baseline and the convergence plane is a way to make the intraocular appear in the right viewing rate. This will require increasing the Toe-in angle for all the cameras which in one hand will improve the multi-view feature and the receiver will be able to see more details from the different viewing angles. In the other hand the stereo feature will be destroyed and the mismatch between the camera’s projection plans will be bigger. Figure 23 illustrates the problem.

"Figure 21a illustrates the projection problem that occurs by using the Toe-in method. Figure 21b illustrates the case when using the Off-Axis method. Figure 22 A screen shot from the display to illustrate the viewing distance"
Figure 23 Left image shows the camera setup with larger mismatch when the convergence plane is closer to the camera’s baseline, while the image on the right shows a smaller mismatch when the convergence plane is further away from the camera’s baseline.

4.7. Data Analysis for the Subjective Evaluation

Twenty-six test subjects participated in the test; they were aged between 19-37 years, 10 of whom were female and 16 male. The subjects had different nationalities and different backgrounds. Most of them do watch “3D-movies” few times a year, and some have never done it before. None of them have ever experienced an autostereoscopic display.

4.7.1. Questionnaire

The Mean Score Opinion for each question is compared between the two systems, and all the questions are listed in the Table 5 below.

Table 5 Comparison for the mean value for each question between the two systems

<table>
<thead>
<tr>
<th>Distance between baseline and convergence plane (cm)</th>
<th>MOS 2D Q1</th>
<th>MOS 3D Q1</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>3.5</td>
<td>3.0</td>
<td>0.040</td>
</tr>
<tr>
<td>1.0</td>
<td>3.2</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>2.9</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>2.7</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>2.5</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>2.3</td>
<td>1.9</td>
<td></td>
</tr>
</tbody>
</table>

The monoscopic system has scored a higher MOS value for the overall audio-visual quality. The p-value is lower than the critical value and the results are significantly different. These results were expected since the monoscopic system delivers a higher image quality with no delay.
The monoscopic system has again scored a higher MOS value here. The p-value is lower than the critical value and that presents a significant difference between the two columns. However the difference between the two systems is much smaller than expected when considering the image quality presented in each of them.

The autostereoscopic system has scored better results here and has a considerable higher value than the monoscopic system. The fact that the p-value is equal to zero shows a definite significant difference. The error bars shows less spreading of score for the autostereoscopic system. The main reason is the Depth-based tasks which gave the subjects a better understanding of the missing depth information in two-dimensional presentation.

The monoscopic system has registered a higher value for the reality of virtual representation of the other person. The fact that the p-value is lower than the critical value confirms that these results are significantly different. This question has presented unexpected results since the autostereoscopic system is supposed to present a higher value which is not the case here. By looking at the comments registered for the test subjects it seems that the reason for the result is the delay that occurred in the 3D system, which has caused lack of reality of the virtual representation of the other person. Another reason could be that the test subjects have experienced the virtual representation of the other person as unusual compared to what they used to see in two-dimensional displays and therefore they scored it with lower value.
The results here are very equal, the p-value is bigger than the critical value and the error bars show a huge overlapping, therefore the results are definitely not statistically different. The interesting part here is that the autostereoscopic systems has scored almost equal MOS value to the monoscopic system despite the delay in the video image.

4.7.2. Ball-Drop Task

The three different cases are presented separately in Table 6 here,

Table 6 Results from the Ball-Drop Task

The overall mean shows a significant improvement in the performance for the subjects while using the autostereoscopic system, the error bars also show less of a spread of scores compared to the monoscopic system.

The average for each field shows that the middle field in the table had always the highest values compared to the two other fields. The errors bars show that the subjects had a smaller error in each field while using the autostereoscopic system compared to the monoscopic system where the spreading of scores is large.
The three averages for the three tries is very close while using the autostereoscopic system, and that confirms that the subjects were well-aware about the position of the ball instead of just taking a chance which is the case for the monoscopic system where averages are different.

By looking at the result for each subject separately in figure 24, it seems that all the subjects had scored higher values by using the autostereoscopic system, by mean depth information brings additional benefits for videoconferencing system and especially in cases where the depth perception is important for the user.

![Figure 24](image)

Figure 24 The mean value of ball score compared between the two systems for each test person, where 2D presents the monoscopic system and 3D presents the autostereoscopic system.

To confirm that the results for the Ball-Drop task are significantly different between the two systems an additional statistic test were performed. The test is known as ANOVA, Analysis of Variance (Maxwell, S. E. and Delaney, H. D. 2003) [14] and recommended by ITU [10]. The results of the ANOVA test are presented in Appendix VII.
4.7.3. Judge Distance Task

The average value for number of correct guesses has been calculated Figure 24 presents the results.

![Figure 24](image)

*Figure 24* The judge distance task for the two systems

The average of correct guesses while using the autostereoscopic system is larger than in the other case while using the monoscopic system. That confirms that the subjects had a better comprehension of the depth in the image presented for them by the autostereoscopic system. The multi-view feature was also a tool to see a better angle of the whiteboard which made it easier to judge the distance between the magnets. The error bars shows an equal spreading of scores for both systems.

4.7.4. Answers and Comments

Most of the test subjects have pointed out the delay in the video image in the autostereoscopic system as disturbing, and that it is mostly important to be improved. The video quality has also been commented as bad and blurry which is understandable in the view of the low resolution that has been used.

On the other hand, all of the subjects described the experience of the autostereoscopic system as new and interesting, and they have also described the presentation of the depth in the image and the multi-view feature as very good and “cool”.

The test subjects were asked if they think that 3D brings any added value to the videoconferencing, and they have been given three alternatives, Yes, No and I do not know. Their answers are illustrated in Figure 26.
Finally, the test subjects were also asked which system has provided them with the best audio-visual quality, and 65% of the total number of subjects chose the monoscopic system as a better system for providing a better audio-visual quality, which is illustrated in Figure 27.

All of the subjects agreed that the video quality and audio-visual quality need to be improved for the autostereoscopic system, and the viewing distance is too large for having a feeling of presence.
5. Conclusion

Several methods have been investigated for generating eight views that can be used for the videoconference system. Using a depth camera can be possible with a better device that can provide a depth map with high resolution. A high resolution depth map can be processed afterwards and undefined areas can be filled using different methods. Using several depth cameras is also an option but that will make the system complicated and at that point using eight real cameras will be a better option. The advantage with view synthesis is less number of images transferred through the network and the view synthesis will be carried out on the receiver’s side. However, the progress is complicated and consists of different steps, which need to be repeated for each view and for each frame. Therefore, a fast hardware system is required to guarantee negligible delay for the signal and no desynchronizing between the different views. An eight camera system can be implemented assuming that enough bandwidth will be provided, the purpose of which is to have results that can be used as a reference for the development of any other enhancing methods in future work.

The subjective testing has provided an overall view that the autostereoscopic system still needs to be improved in several aspects, such as video quality. Sense of depth and multi-view features are provided by the system and give the intraocular better understanding of the environment where the other party is located. However, the need of these features depends on what field of application the videoconference system is being used for. For example, the sense of depth might not be necessary when having a normal conversation between two people, while it could be important for a physiotherapist to see at what angle the patient is moving his arm related to the rest of his body.
6. Future Work

Future work can be carried on in different areas for improving the system. There are novel technologies for autostereoscopic displays with shorter viewing distance as much as two metres, and that could improve the sense of presence. The new multi-camera setup needs to be designed with better cameras that deliver identical quality of images and have user-controlled parameters such as focus, field of view and zoom lens which make it possible to calibrate the cameras. The problem of delay in the video image needs to be sorted out by optimizing the rendering algorithm on the autostereoscopic display, and a Multi-view coding (MVC) might be used to encode the different streams captured by the different cameras.

However, new autostereoscopic displays will be available with at least 24 views within few years, and using real cameras for capturing this amount of different views is not the ideal solution. Therefore an optimized algorithm for view-synthesize needs to be developed to reduce the number of cameras and the capacity needed for the network.

A new subjective test can be performed with the new conditions and more controlled parameters to compare it to the one that has been performed in this work and see if any applied enhancing methods will make a change in the results for the test.

The development of different technologies is increasing for each day and at a fast pace. Autostereoscopic displays might replace traditional displays in every home within a few years. Therefore, it will not be surprising if autostereoscopic videoconferencing will be common in the near future even for private usage where the need of a sense of presence will be more than usual in video communication.
Reference List


Appendix I

Terms & Definitions

A. 

Accommodation - This is an oculomotor cue for depth perception. When we try to focus on far away objects, the ciliary muscles stretch the eye lens, making it thinner. The kinesthetic sensations of the contracting and relaxing ciliary muscles (intraocular muscles) are sent to the visual cortex where it is used for interpreting distance/depth.

B. 

Binocular vision is vision in which both eyes are used together.

Binocular disparity refers to the difference in image location of an object seen by the left and right eyes, resulting from the eyes' horizontal separation. The brain uses binocular disparity to extract depth information from the two-dimensional retinal images in stereopsis.

Binocular rivalry is a phenomenon of visual perception in which perception alternates between different images presented to each eye.

C. 

Camera convergence angle in the toed-in camera model, this is the angle the cameras are each rotated inwards from parallel to achieve convergence.

Camera field of view is the horizontal angle of view of the camera.

Contextual factor is a factor which involves the context within which the stereo display is viewed. For example the spatio-temporal frequency effects and the distance scaling of disparity information are two contextual factors.

Cross-Talk: In stereoscopic 3D displays, "crosstalk" refers to the incomplete isolation of the left and right image channels so that one leaks or bleeds into the other - like a double exposure, which produces a ghosting effect.

Convergence is when both eyes converge on the object of interest.

Convergence distance is the distance at which the optical axes of the two cameras intersect.
D.

**Depth perception** is the visual ability to perceive the world in three dimensions (3D) and the distance of an object.

**Dipoter** is a unit used to measure the diffractions ability of lenses, (i.e. the depth of field is calculated using the depth of focus by \( D = 1/F \) where \( D \) is distance in meters and \( F \) is distance in diopters).

**Distance scaling of disparity** hence the disparity information must be re-calibrated for different viewing distances in order to reliable veridical depth to be perceived.

E.

**Egocentric viewing distance** refers to the distance between an observer and a given fixation object out in the visual field.

**Epipole or epipolar point** Since the centres of projection of the cameras are distinct, each centre of projection projects onto a distinct point into the other camera’s image plane. These two image points are denoted by \( e_L \) and \( e_R \) and are called *epipoles* or *epipolar points*.

**Eye separation** is the distance between the observer’s eye, typically 65mm.

F.

**Frame magnification** is the ratio of screen width to camera sensor width.

H.

**Heterophoria** is a type of eye condition in which the directions that the eyes are pointing are not consistent with each other.

**Holography** is a technique which enables three-dimensional images to be made. It involves the use of a laser, interference, diffraction, light intensity recording and suitable illumination of the recording. The image changes as the position and orientation of the viewing system changes in exactly the same way as if the object were still present, thus making the image appear three-dimensional.

**Horopter** defines the positions in the visual field that also give rise to pairs of images that stimulate corresponding retinal areas in the two eyes and therefore possess zero disparity.

I.

**Image parallax** is the horizontal distance between homologous points on the screen.

**Interlocutor** is a person taking part in a conversation between people via a tele-meeting system.

K.

**Keystone distortion** is an image distortion which causes vertical parallax in stereoscopic video image due to the imaging sensors of the two cameras being located in different planes.
M.

**Moving parallax** is seeing different images when we move our head.

**Monocular cues** provide depth information when viewing a scene with one eye.

P.

**Parallax** is a displacement or difference in the apparent position of an object viewed along two different lines of sight, and is measured by the angle or semi-angle of inclination between those two lines.

Q.

**Quality of experience (QoE)**, sometimes also known as **quality of user experience**, is a subjective measure of a customer's experiences with a service (web browsing, phone call, TV broadcast, call to a Call Center).

R.

**Relative depth** refers to the depth interval between an object and some reference point such as the horopter.

S.

**Screen width** is the horizontal size of the display screen.

**Sensor axial offset** is the distance by which the centre of each imaging sensor (CCD) has been moved away (outwards) from the optical axis of the lens to achieve convergence, in the parallel camera configuration.

**Shear distortion** is different type of distortion caused by sideways movement of the observer.

**Spatio-temporal frequency effects** refers to the rate of modulation of luminance information over space and time in an image.

**Stereoanomaly** is the condition where for some individuals the perceived depth is reversed such that depth introduced by crossed disparity is actually perceived as back depth or depth induced by uncrossed disparity is actually perceived as front depth.

T.

**Telemeeting** is a meeting which participants are located at least two locations and the communication takes place via a telecommunication system.
Vergence is the simultaneous movement of both eyes in opposite directions to obtain or maintain single binocular vision.

Viewing distance is the distance from the observer’s eyes to the display plane.

Visual perception is the ability to interpret the surrounding environment by processing information that is contained in visible light.

Volumetric display device is a graphical display device that forms a visual representation of an object in three physical dimensions.
Appendix II

Table of equipment characteristics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Room 1</th>
<th>Room 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room 1 (Show) size</td>
<td>7x4x2.8</td>
<td>7x4x2.8</td>
</tr>
<tr>
<td>Autostereoscopic Display size</td>
<td>93.5 x 53 cm</td>
<td>93.5 x 53 cm</td>
</tr>
<tr>
<td>Viewing distance*</td>
<td>≈ 4 metres</td>
<td>≈ 4 metres</td>
</tr>
<tr>
<td>Peak Luminance of the screen</td>
<td>168 cd/m2</td>
<td>168 cd/m2</td>
</tr>
<tr>
<td>Ratio of luminance of inactive screen to peak</td>
<td>1:1680</td>
<td>1:1680</td>
</tr>
<tr>
<td>luminance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio of the luminance to the screen when</td>
<td>1:1680</td>
<td>1:1680</td>
</tr>
<tr>
<td>displaying only black level in a completely</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dark room, to that corresponding to peak white</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio of luminance of background behind</td>
<td>1:1680</td>
<td>1:1680</td>
</tr>
<tr>
<td>picture monitor to peak luminance of picture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromacity of background</td>
<td>D65</td>
<td>D65</td>
</tr>
<tr>
<td>Background room illumination</td>
<td>≈ 20 lux</td>
<td>≈ 20 lux</td>
</tr>
<tr>
<td>Background noise level**</td>
<td>&lt;= 30 dBA</td>
<td>&lt;= 30 dBA</td>
</tr>
</tbody>
</table>

* The viewing distance is 4 metres as it is the recommended distance for the autostereoscopic display.

** Since the headphones that are used has a noise cancellation function, that secure low background noise level.
Appendix III

Confero Subjective Test: Session 1

In this session you will be testing monoscopic videoconferencing.

How would you rate the reality of the virtual representation of the other person? *
1) Bad 2) Poor 3) Fair 4) Good 5) Excellent

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bad</td>
</tr>
</tbody>
</table>

How would you rate the overall audio-visual quality? *
1) Bad 2) Poor 3) Fair 4) Good 5) Excellent

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bad</td>
</tr>
</tbody>
</table>

In what grade did you experience that the other party was present in the same room? *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bad</td>
</tr>
</tbody>
</table>

How would you rate the video quality of the connection? *
1) Bad 2) Poor 3) Fair 4) Good 5) Excellent

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bad</td>
</tr>
</tbody>
</table>

How would you rate the quality of depth perception? *
1) Bad 2) Poor 3) Fair 4) Good 5) Excellent

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bad</td>
</tr>
</tbody>
</table>

How would you judge the effort needed to interrupt the other party?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>No effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extreme effort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bad</td>
</tr>
</tbody>
</table>

Did you have any difficulty during the connection? *

- Yes
- No
- I do not know

Was the connection acceptable? *

- Yes
- No
- I do not know
Confero Subjective Test: Session 2

In this session you will be testing multi-channel stereoscopic videoconferencing.

How would you rate the overall audio-visual quality? *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
</table>

Did you have any difficulty during the connection? *

- Yes
- No
- I do not know

How would you rate the video quality of the connection? *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
</table>

How would you rate the quality of depth perception? *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
</table>

In what grade did you experience that the other party was present in the same room? *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
</table>

Do you think 3D can bring added value to the videoconference? *

- Yes
- No
- I do not know

How would you rate the reality of the virtual representation of the other person? *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Excellent</th>
</tr>
</thead>
</table>

How would you judge the effort needed to interrupt the other party? *

<table>
<thead>
<tr>
<th>Extreme effort</th>
<th>No effort</th>
</tr>
</thead>
</table>

Was the connection acceptable? *

- Yes
Confero Subjective Test: Post-Screen

Since we are at the end of the test, I would like to ask you some supplementary questions. Are you ready?

Which mode provided you with the best overall audio-visual quality? *

- [ ] Monosocopic (2D) one-view
- [ ] Stereoscopic (3D) multi-view

Could you tell me, in order of importance, which of the following factors would you consider sufficiently important that you would like to see improved for the 3D system. (Optional)
(1) is very important and (5) is not important

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio-visual Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth Perception</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do you have any other comments relative to the entire test you wish to make known to us? (Optional)

Would you like to participate in any future tests? (Optional)
If you press yes, we will keep your Email and contact you in future.

- [ ] Yes
- [ ] No
Appendix IV

Table of scores for questions Q1-Q5

<table>
<thead>
<tr>
<th>Test-ID</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Gender</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>3027101</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>M</td>
<td>25</td>
</tr>
<tr>
<td>9412259</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>M</td>
<td>22</td>
</tr>
<tr>
<td>5298264</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>M</td>
<td>21</td>
</tr>
<tr>
<td>6693388</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>M</td>
<td>29</td>
</tr>
<tr>
<td>7951337</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>M</td>
<td>29</td>
</tr>
<tr>
<td>4977905</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>F</td>
<td>27</td>
</tr>
<tr>
<td>7043593</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>F</td>
<td>30</td>
</tr>
<tr>
<td>8326691</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>M</td>
<td>24</td>
</tr>
<tr>
<td>5654211</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>M</td>
<td>16</td>
</tr>
<tr>
<td>7739787</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>F</td>
<td>18</td>
</tr>
<tr>
<td>6604296</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>M</td>
<td>24</td>
</tr>
<tr>
<td>7153084</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>F</td>
<td>30</td>
</tr>
<tr>
<td>6011488</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>F</td>
<td>30</td>
</tr>
<tr>
<td>9305422</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>M</td>
<td>26</td>
</tr>
<tr>
<td>7716220</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>M</td>
<td>31</td>
</tr>
<tr>
<td>1083108</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>M</td>
<td>37</td>
</tr>
<tr>
<td>1424115</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>M</td>
<td>32</td>
</tr>
<tr>
<td>6066291</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>M</td>
<td>19</td>
</tr>
<tr>
<td>4189893</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>M</td>
<td>26</td>
</tr>
<tr>
<td>4971445</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>M</td>
<td>27</td>
</tr>
<tr>
<td>8715698</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>F</td>
<td>26</td>
</tr>
<tr>
<td>2405804</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>F</td>
<td>25</td>
</tr>
<tr>
<td>2233620</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>F</td>
<td>21</td>
</tr>
<tr>
<td>3737281</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>M</td>
<td>20</td>
</tr>
<tr>
<td>9165804</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>F</td>
<td>24</td>
</tr>
<tr>
<td>4083719</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>F</td>
<td>23</td>
</tr>
</tbody>
</table>
## Appendix V

### Table of scores for Ball-Drop Task

<table>
<thead>
<tr>
<th>Test-ID/Question</th>
<th>Monoscopic</th>
<th>Autostereoscopic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 1 2 3</td>
<td>1 2 3 1 2 3</td>
</tr>
<tr>
<td>3027101</td>
<td>6 8 6 8 6 7 8 8 8</td>
<td>8 8 8 8 8 6 8 8 8</td>
</tr>
<tr>
<td>9412259</td>
<td>8 7 8 7 7 7 7 7 7</td>
<td>7 7 8 8 8 8 7 8 8</td>
</tr>
<tr>
<td>5298264</td>
<td>5 5 6 8 8 5 8 5 5</td>
<td>6 8 8 6 8 8 8 7 8</td>
</tr>
<tr>
<td>6693388</td>
<td>6 6 5 7 5 6 7 6 6</td>
<td>7 8 7 8 8 8 8 8 8</td>
</tr>
<tr>
<td>7951337</td>
<td>6 7 8 6 4 5 8 6 6</td>
<td>8 7 8 8 8 8 8 8 8</td>
</tr>
<tr>
<td>4977905</td>
<td>7 8 7 7 6 7 6 4 6</td>
<td>7 8 7 8 7 7 6 7 6</td>
</tr>
<tr>
<td>7043593</td>
<td>6 5 5 6 8 8 6 8 8</td>
<td>6 8 8 8 8 7 6 8 8</td>
</tr>
<tr>
<td>8326691</td>
<td>8 6 8 6 7 6 6 8 6</td>
<td>6 8 5 8 7 8 6 7 8</td>
</tr>
<tr>
<td>5654211</td>
<td>6 8 8 8 6 8 8 6 6</td>
<td>8 8 7 8 8 8 8 8 8</td>
</tr>
<tr>
<td>7739787</td>
<td>6 6 8 8 8 8 6 5 5</td>
<td>7 8 8 8 8 8 8 8 8</td>
</tr>
<tr>
<td>6604296</td>
<td>7 7 6 8 8 6 8 8 5</td>
<td>8 8 8 6 7 8 8 7 6</td>
</tr>
<tr>
<td>7153084</td>
<td>6 4 8 8 8 6 8 5 8</td>
<td>8 8 8 8 6 8 8 8 6</td>
</tr>
<tr>
<td>6011488</td>
<td>8 6 8 6 5 8 6 6 8</td>
<td>7 8 8 8 8 8 8 7 8</td>
</tr>
<tr>
<td>9305422</td>
<td>6 6 8 8 7 8 6 6 8</td>
<td>8 8 8 8 8 8 7 8 8</td>
</tr>
<tr>
<td>7716220</td>
<td>6 6 8 5 6 6 6 6 6</td>
<td>8 8 7 8 8 8 6 8 8</td>
</tr>
<tr>
<td>1083108</td>
<td>6 6 8 8 6 8 6 3 8</td>
<td>8 7 7 8 7 8 8 8 7</td>
</tr>
<tr>
<td>1424115</td>
<td>5 6 6 6 6 6 6 8 6</td>
<td>8 7 8 8 8 8 8 8 8</td>
</tr>
<tr>
<td>6066291</td>
<td>6 6 6 6 6 6 6 6 5</td>
<td>6 8 8 8 8 8 8 8 8</td>
</tr>
<tr>
<td>4189893</td>
<td>6 4 6 6 6 7 6 8 6</td>
<td>8 7 8 8 7 8 6 7 7</td>
</tr>
<tr>
<td>4971445</td>
<td>4 8 6 6 6 8 7 7 6</td>
<td>7 6 6 7 7 8 6 8 7</td>
</tr>
<tr>
<td>8715698</td>
<td>6 7 8 8 6 6 7 7 8</td>
<td>8 7 6 7 8 8 6 8 6</td>
</tr>
<tr>
<td>2405804</td>
<td>8 6 8 5 7 8 6 6 4</td>
<td>6 8 8 8 8 7 8 8 8</td>
</tr>
<tr>
<td>2233620</td>
<td>7 8 6 4 7 5 6 8 6</td>
<td>7 8 8 8 8 8 8 7 6</td>
</tr>
<tr>
<td>3737281</td>
<td>5 5 7 5 6 8 6 8 7</td>
<td>7 7 8 8 8 8 8 8 7</td>
</tr>
<tr>
<td>9165804</td>
<td>7 8 6 6 6 6 8 6 7</td>
<td>8 8 7 8 6 7 8 7 8</td>
</tr>
<tr>
<td>4083719</td>
<td>6 5 8 8 6 7 8 8 8</td>
<td>6 8 6 8 8 8 8 8 8</td>
</tr>
</tbody>
</table>
## Appendix VI

### Table of scores for Judge-Distance Task

<table>
<thead>
<tr>
<th>Test-ID</th>
<th>Monoscopic</th>
<th>Autostereoscopic</th>
</tr>
</thead>
<tbody>
<tr>
<td>3027101</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9412259</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5298264</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>6693388</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>7951337</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4977905</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7043593</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>8326691</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5654211</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>7739787</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6604296</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7153084</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6011488</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>9305422</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7716220</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1083108</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1424115</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6066291</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4189893</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>4971445</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8715698</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2405804</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2233620</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>3737281</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>9165804</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4083719</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Appendix VII

ANOVA Test of Ball-Drop Task

<table>
<thead>
<tr>
<th>Main effect</th>
<th>2D/3D</th>
<th>Mean</th>
<th>Std err</th>
<th>-95% conf</th>
<th>+95 conf</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D vs 3D</td>
<td>2D</td>
<td>1</td>
<td>1</td>
<td>6.572650</td>
<td>0.067613</td>
<td>6.436844 6.708455</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>2</td>
<td>2</td>
<td>7.555556</td>
<td>0.067613</td>
<td>7.419750 7.691361</td>
</tr>
</tbody>
</table>

![2D vs 3D](image)

<table>
<thead>
<tr>
<th>Interaction 2D/3D and Field</th>
<th>2D/3D</th>
<th>Field</th>
<th>Mean</th>
<th>Std err</th>
<th>-95% conf</th>
<th>+95 conf</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2D</td>
<td>1</td>
<td>1</td>
<td>6.512821</td>
<td>0.118783</td>
<td>6.274240 6.751401</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>2D</td>
<td>2</td>
<td>2</td>
<td>6.615385</td>
<td>0.113242</td>
<td>6.387931 6.842838</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>2D</td>
<td>3</td>
<td>3</td>
<td>6.589744</td>
<td>0.116392</td>
<td>6.355964 6.823523</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>1</td>
<td>1</td>
<td>7.410256</td>
<td>0.118783</td>
<td>7.171676 7.648837</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>2</td>
<td>2</td>
<td>7.705128</td>
<td>0.113242</td>
<td>7.477675 7.932582</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>3</td>
<td>3</td>
<td>7.551282</td>
<td>0.116392</td>
<td>7.317503 7.785061</td>
<td>26</td>
</tr>
</tbody>
</table>
### Interaction 2D/3D and Try

<table>
<thead>
<tr>
<th>2D/3D</th>
<th>REPS</th>
<th>Mean</th>
<th>Std err</th>
<th>-95% conf</th>
<th>+95% conf</th>
<th>N</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>6.538462</td>
<td>0.112660</td>
<td>6.312177</td>
<td>6.764746</td>
<td>26</td>
<td>0.226285</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>6.448718</td>
<td>0.104359</td>
<td>6.239106</td>
<td>6.658330</td>
<td>26</td>
<td>0.209612</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>6.730769</td>
<td>0.118810</td>
<td>6.492133</td>
<td>6.969405</td>
<td>26</td>
<td>0.238636</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>7.525641</td>
<td>0.112660</td>
<td>7.299356</td>
<td>7.751926</td>
<td>26</td>
<td>0.226285</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>7.564103</td>
<td>0.104359</td>
<td>7.354491</td>
<td>7.773714</td>
<td>26</td>
<td>0.209612</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>7.576923</td>
<td>0.118810</td>
<td>7.338287</td>
<td>7.815559</td>
<td>26</td>
<td>0.238636</td>
</tr>
</tbody>
</table>

### Repeated Measures Analysis of Variance (Ball-Drop Task)
Sigma-restricted parameterization
Effective hypothesis decomposition; Std. Error of Estimate: 1.0342

<table>
<thead>
<tr>
<th>Effect</th>
<th>SS</th>
<th>Degr. Of Freedom</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>23353.92</td>
<td>1</td>
<td>23353.92</td>
<td>21831.33</td>
<td>0.000000</td>
</tr>
<tr>
<td>2D/3D</td>
<td>113.03</td>
<td>1</td>
<td>113.03</td>
<td>105.66</td>
<td>0.000000</td>
</tr>
<tr>
<td>Error</td>
<td>53.49</td>
<td>50</td>
<td>1.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARGET</td>
<td>3.09</td>
<td>2</td>
<td>1.54</td>
<td>1.48</td>
<td>0.232585</td>
</tr>
<tr>
<td>TARGET*2D/3D</td>
<td>0.75</td>
<td>2</td>
<td>0.37</td>
<td>0.36</td>
<td>0.699809</td>
</tr>
<tr>
<td>Error</td>
<td>104.38</td>
<td>100</td>
<td>1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRY</td>
<td>1.94</td>
<td>2</td>
<td>0.97</td>
<td>1.03</td>
<td>0.359065</td>
</tr>
<tr>
<td>TRY*2D/3D</td>
<td>1.41</td>
<td>2</td>
<td>0.71</td>
<td>0.76</td>
<td>0.359065</td>
</tr>
<tr>
<td>Error</td>
<td>93.54</td>
<td>100</td>
<td>0.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARGET<em>TRY</em>2D/3D</td>
<td>3.42</td>
<td>4</td>
<td>0.85</td>
<td>1.14</td>
<td>0.340573</td>
</tr>
<tr>
<td>Error</td>
<td>150.44</td>
<td>200</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>