NTNU – Trondheim
Norwegian University of
Science and Technology

Subjective and Objective Crosstalk Assessment Methodologies for Auto-stereoscopic Displays

Kim Daniel Skildheim

Master of Science in Communication Technology
Submission date: June 2012
Supervisor: Andrew Perkis, IET
Co-supervisor: Liyuan Xing, IET

Norwegian University of Science and Technology
Department of Electronics and Telecommunications

ABSTRACT

Stereoscopic perception is achievable when the observer sees a scene from a slightly different angle. Auto-stereoscopic displays utilize several separate views to achieve this without using any special glasses. Crosstalk is an undesired effect of separating views. It is one of the most annoying artefacts occurring in an auto-stereoscopic display. This experiment has two parts. The first part proposes a subjective *assessment* methodology for characterizing crosstalk in an auto-stereoscopic display, *without restriction of subjects' viewing behaviour*. The intention was to create an inexpensive method. The measurement was performed by using a Kinect prime sensor as a head tracking system combined with subjective score evaluation to get a data plot of the perceived crosstalk. The crosstalk varies in line with image content, disparity and viewing position. The result is a data plot that approaches a periodically pattern, which is consistent with the characteristics of an auto-stereoscopic display. The result is not perfect since there are many sources of errors occurring. These errors can be improved with better head tracking, an improved movement system, post processing of data, more data and removal of outliers.

The second part proposes methods for extracting subjective values based on interpolated plots and creating objective crosstalk influenced pictures which correlate with the subjective data. The best extraction method was to combine an adapted sine regression curve with a linear interpolation. This interpolation followed the subjective values in a parallel slice plot at 3.592 m from the screen. The interpolation was adapted to fit a derived model as best as possible to achieve a good correlation. Objective crosstalk pictures were created, where the amount of crosstalk was determined by the neighbouring view that influenced the current view the most. The correlation was based on the relationship between the SSIM value from the created crosstalk picture and the extracted subjective value. The total correlation of the pictures together were 0,8249, where the picture with the highest correlation had 0,9561. This method is quite good for pictures that have a maximum disparity grade below 38 pixels. The overall result is good and it is also a measure of quality for the subjective test. This result can be improved by increasing the complexity of how the objective crosstalk pictures are created by adding more views into account or try another method to create crosstalk. Improved extraction of subjective values will also be beneficial in terms of improving the correlation even more.

Keywords: Crosstalk, views, disparity, subjective methodology, creating objective crosstalk influenced pictures, interpolation, correlation

SAMMENDRAG

En stereoskopisk persepsjon er mulig dersom man observerer et object fra forskjellige vinkler, altså man oppnår en dybdefølelse. I dag bruker man som regel spesialbriller for å separere to bilder og dermed oppleve en 3D effekt på en flat skjerm. En auto-stereoskopisk skjerm sender ut flere bilder 'views' med litt forskjellig vinkel til observatøren. Dersom man ser disse litt forskjellige viewsene vil man oppleve en 3D effekt uten å bruke spesialbriller. En vanlig forstyrrelse som oppstår med denne teknologien kalles 'crosstalk'. Dette er en av de mest irriterende forstyrrelsene som oppstår ved bruk av brillefri 3D.

Denne rapporten er todelt der første del utleder en subjectiv diagnostiseringsmetodikk for å karakterisere crosstalk av en auto-stereoskopisk skjerm, uten å begrense hvor observanten skal posisjonere seg. En av intensjonene var at metoden skulle appellere til et lavt budsjett. Posisjonen til observanten blir målt ved hjelp av en Kinct Prime sensor som bruker en hodegjenkjennings-algoritme for å finne hode-posisjonen. Observanten vurderer 3D bildet fra flere posisjoner og gir en verdi som er avhengig av mengden crosstalk som forstyrrer opplevelsen av bilde. Crosstalk varierer i forhold til innhold i bilde, avstand mellom hvert bilde som sendes ut og posisjon. Resultatet ble et periodisk mønster som var konsistent med en auto-stereoskopisk skjerm og det man forventet teoretisk. Resultatet ble ikke helt perfekt siden det er vanskelig å neglisjere alle feil og forstyrrelser som kan oppstå. Testen kan forbedres med en bedre sensor for å finne posisjonen til observanten, bedre bevegelses system for observanten, pre-prosessering av data, mere data og bedre forståelse for å finne observanter i testen som bør fjernes fra resultatet på grunn av store avvik fra andre observanter.

Del to av raporten foreslår metoder for ekstrahering av subjektive data som er basert på interpolerte metoder og hvordan lage crosstalk-påvirkede bilder objektivt som skal kunne representere det observanten observerte i den subjektive testen med en god korrelasjon. Den beste ekstraherings-metoden var å kombinere en sinus regresjonskurve sammen med en lineær interpolasjon kurve. Denne interpolasjonskurven ble brukt til å punktprøve subjektive verdier fra resultatet i den subjektive testen. En adaptiv sinus regesjonskurve kombinert med den lineære kurven i et paralellt plan med skjermen på en avstand ved 3.592m gav det beste resultatet. Crosstalk-påvirkede bilder ble laget basert på mengden crosstalk som kommer fra det nærliggende view som påvirket mest. Korrelasjonen ble basert på forholdet mellom SSIM verdier og den tilhørende ekstraherte subjektive verdien. Korrelasjonen for alle bildene ble på 0.8249 der beste bilde hadden en korrelasjon på 0.9561. Metodene viser seg å være gode for bilder som har en max ulikhet mellom viewene på 38 pixler. Dette er et godt resultat og avspeiler også kvaliteten av den subjective testen. Resultatet kan forbedres ved å øke kompleksiteten rundt hvordan man lager de objektive bildene og forbedre ekstrasjonen av subjektive verdien ved hjelp av en bedre interpolasjon.

Table of contents

1: INTRODUCTION

Auto-stereoscopic displays that utilizes slanted lenticular sheets, produces different artefacts. These artefacts influence the received 3D perception. One of these artefacts is called crosstalk, and is caused by a process which can be modelled as inter-channel crosstalk [3].The overall goal of the project is to*: "Propose a subjective crosstalk assessment methodology that is realistic without restriction of subjects' viewing behaviour and create objective crosstalk influenced pictures to see if they correlate with the subject results based on the derived methodology".*

It is not performed a subjective assessment that do not restrict the subjects' viewing behaviour as a study before. The result can be used to calculate a metric to characterize auto-stereoscopic displays. The experiment performed was divided into two parts, first the subjective test part and second the objective part. This report will go through each step of the process starting with theory and will continue with test setup and implementation, results and discussion. This experiment proposes a methodology for subjective crosstalk measurements and creation of objective crosstalk influenced pictures. The correlation between the subjective and the objective results will be a measure of the quality of the experiment. The proposed methodology was designed and carried out by Liyuan Xing, Jie Xu and Kim Daniel Skildheim.

2: THEORY

This section will derive theory about auto-stereoscopic technology, crosstalk, the SSIM algorithm and basic Pearson correlation. It forms the fundamental basis of what one can expect of the test setup, results and discussion.

SUBJECTIVE TESTING

2.1 AUTO-STERIOSCOPIC TECHNOLOGY

Stereoscopic 3D perception is achievable when each eye of the observer sees a scene from different angles [3]. A display that creates this effect without requiring the observer to use special glasses like anaglyphic lenses, polarized glasses or liquid crystal shutter glasses, are called an auto-stereoscopic display [1]. There are several ways to realize this illusion with different technologies [3]. Multiple projectors, lenticular lenslets, parallax barriers, Fourier-plane shuttering, retro-reflective mirrors and half-silvered mirrors are some of the most common technologies that can achieve the 3D-illusion [5].By illusion we try to mimic the behaviour of the human visual system by using stereo cameras to capture a scene from slightly different positions. Just like our eyes sees a scene from a slightly different position. Viewing a real world scene, there are an infinite number of possible images of the scene. It is possible to conceptually divide this viewing space into a finite number of *views* [6].

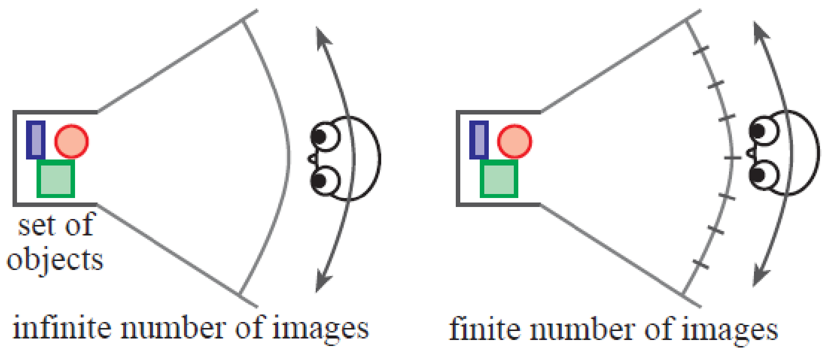


Figure 1. Infinite versus finite number of images [6]

The most popular multiview 3D displays are lenticular/spherical lenses and parallax barriers based screens [13]. In this experiment there was used a multiview Tridelity MV4200 3D TV, which utilizes lenticular lenses. The display works by simultaneously showing a set of images (views). This is done by splitting the native resolution into a number of views [1]. These views can be seen from a number particular viewing angles, which is oriented along the horizontal axis. Figure 2 shows a 5 view distribution [3]. The views are separated by adding optical filters that adjusts the propagation of the current view. The optical filters can vary from different vendors that produce multiview screens. The result of this is that the design parameters are different and will give various tradeoffs between screen resolution, number of views and optimal viewing distance [13]. The numbers of views vary from different technology and vendors operate with diverse number of views on their multiview screens. The most common are from 5 to 15 views [5].

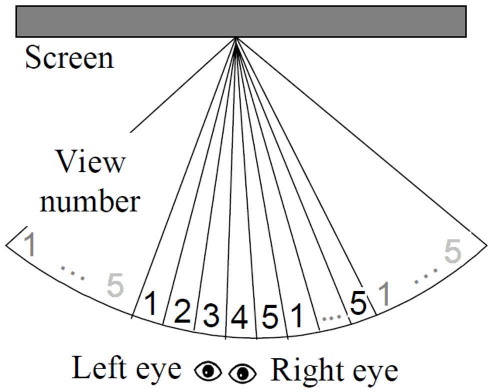


Figure 2. 5 view distribution [3, refined]

Each view has its own set of sub-pixels that correspond to the specific view. In this way the observer is able to see two different views, one view for each eye. This makes it possible to get the stereoscopic perception. Each view can also be seen from a number of observation positions, since the whole set of views are repeated along the horizontal axis. When the observer moves in the transition area between the last view in a set of views and to the next set, in this case between view 5 and view 1, it produces a characteristic *jump* for the observer's perception. It can be perceived that the image distorting [3]. In the rest of the report is view 1 similar to V1, view 2 similar to V2 etc.

2.2 CROSSTALK

In order to crate the 3D illusion, the slanted lens array technology is one of the most popular approaches, which is used in this experiment as well. The slanted lens array is used to balance the horizontal versus the vertical resolution of the display [2]. It causes the sub- pixels of a view to appear on a non-rectangular grid as illustrated in figure 3[3].

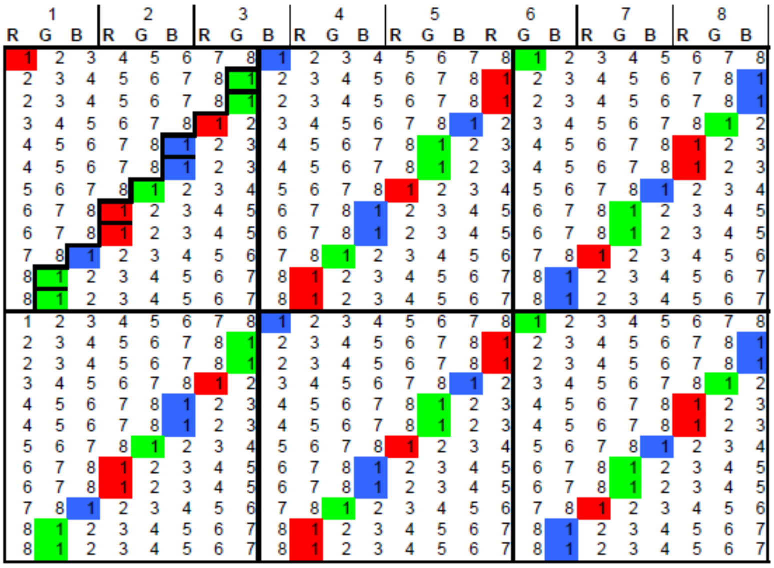


Figure 3. Sub-pixels on a non-rectangular grid[3]

This technique requires specially designed filters to prevent aliasing caused by the sub-sampling of this grid[9,10]. Another effect which is much more distinct and annoying, ***is that parts of the neighbouring views are seen in the current view. This is called crosstalk.*** This artefact is perceived as "double edges" or ghosting along objects in the image, which is illustrated in figure 4. Since the lenticular lenses are slanted, the sub-pixels' boundaries cannot be covered by the lens elements exactly. The outcome is that some sub-pixels appear only partly in the current view. In addition, the centre of the lens element will appear arbitrary displaced over a sub-pixel triplet. The result of this is that the optimal observation point will be somewhat shifted for different sub-pixels of a certain view. Different vendors creates a more uniform view by broaden the observation angle due to compensate to this arbitrary placement. All these effects affect the sub-pixels from neighbouring views to be cast towards the current view and this causes crosstalk [3].

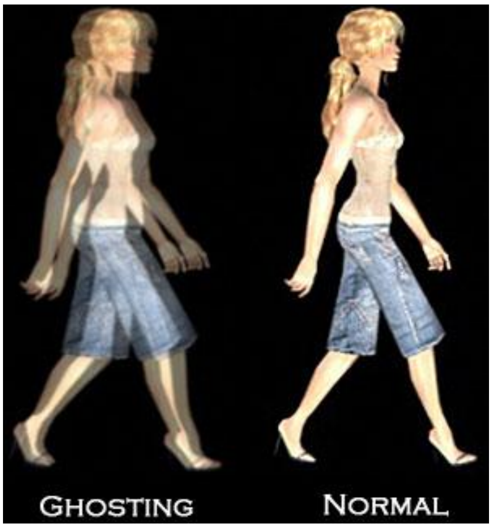


Figure 4. Double edges or ghosting [27]

The amount of crosstalk depends on several production parameters, like the design of the lenticular sheet, the pixels and the placement of the sheet related to the screen and the distance between the sheet and the pixels. Given a picture, the largest disparity between neighbouring views will certainly influence the amount of crosstalk as well. System crosstalk in auto-stereoscopic displays has three different features which distinguish it from a stereoscopic display: 1) the crosstalk depends on the observation position; 2) Most crosstalk appear from the neighbour views and some from the rest of the views; 3) The amount of crosstalk from the neighbour views varies along the horizontal and vertical axes of the screen[1] and 4) the crosstalk perceived is dependent of the picture content. Moreover, some crosstalk is argued to have benefit in multi-view auto-stereoscopic displays in [8], by accounting for optical crosstalk leads to increased bandwidth of the ideal anti-alias filter.

2.3 STRUCTURAL SIMILARITY INDEX (SSIM)

The Structural SIMilarity (SSIM) index is a method for measuring the similarity between two images. It can be viewed as a quality measure of one of the images being compared, provided the other image is regarded as of perfect quality. The distorted image is compared to an original image of perfect quality. The algorithm is an improved version of the universal image quality index [12].

The SSIM is designed to improve the methods peak signal-to-noise ratio (PSNR) and mean squared error (MSE), which have been proved to be inconsistent with human perceptions of images and video [18]. The SSIM evaluate image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. The SSIM index are calculated of by divide picture X and Y into windows, often the size of 8x8 pixels. A SSIM value is calculated for each window by the equation 1,

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \qquad (1)$$

where μ_x and μ_y are the average of x and y, $\sigma^2{}_x$ and $\sigma^2{}_y$ are the variance of x and y, σ_{xy} is the covariance of x and y and $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ are two variables that stabilizes the division with weak denominator. L is the dynamic range of the pixel values ($2^{\#\text{bits per pixel}}$ - 1) and $k_1 = 0.01$, $k_2 = 0.03$ by default.

The output value has a range from -1 to 1 where the value 1 is the comparison of two identical pictures.

Crosstalk is more perceivable in the structure of the picture rather than the overall quality. This is an important argument that SSIM would give a better measure of crosstalk compared to PSNR or MSE. Some type of distortions will not be detected by the MSE although it is easy to perceive by the human perception. The SSIM algorithm will often detect additional types of distortion compared to MSE [17]. Figure 5 gives an example. The MSE value is nearly constant at 144 while the SSIM value varies from 1 to 0.662

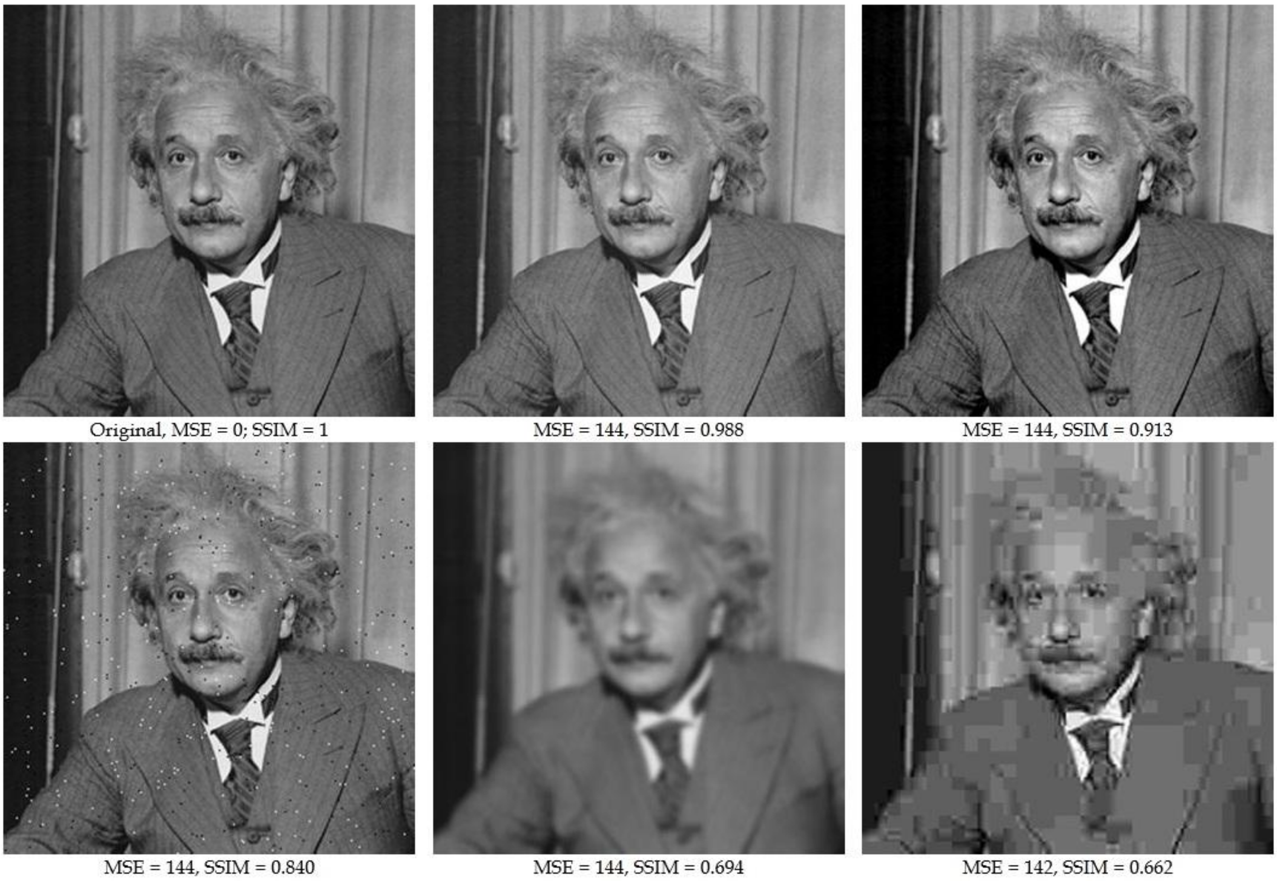


Figure 5. Comparison between MSE and SSIM [17]

A SSIM map unveils where the structural dissimilarity appears. This map gives a good indication of dissimilarity especially in crosstalk distorted images. Figure 6 below compare a high resolution image with a JPEG compressed image with a SSIM map and Absolute error map.

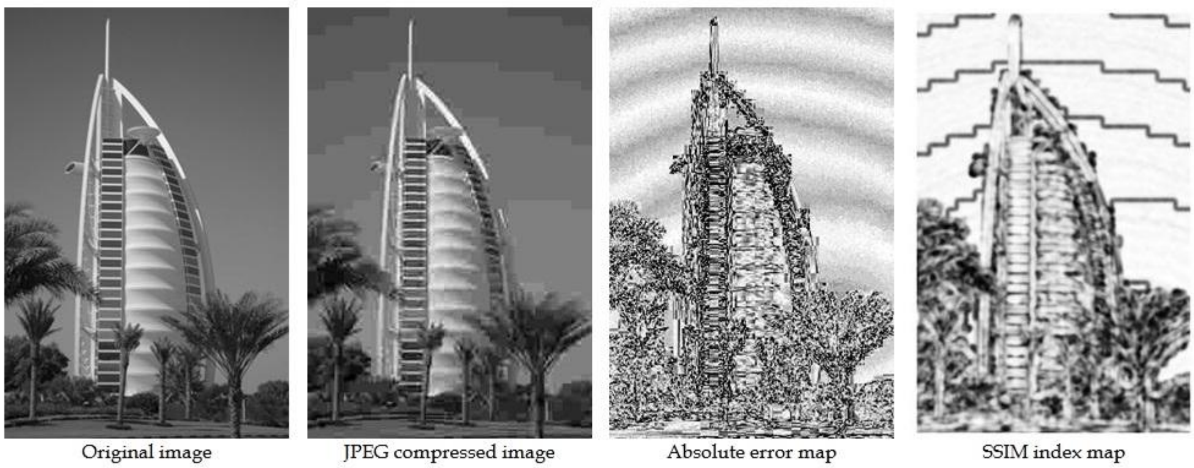


Figure 6. Image comparison between original, JPEG compressed, error map and SSIM map.

2.4 SUBJECTIVE TESTING

This section will derive theory about subjective testing, concepts and methodology.

It is not possible to fully characterize system performance by objective means; consequently, it is necessary to supplement objective measurements with subjective measurements. In general there are two classes of subjective assessments. The first one is *quality assessments* that establish the performance of systems under optimum conditions. The second is called *impairment assessments* that establish the ability of systems to retain quality under non-optimum conditions that relate to transmission or emission. A subjective test has to be reproducible, thus the experimenter has to provide full description of the laboratory environment, test material, observers, test method and who to process the subjective data.

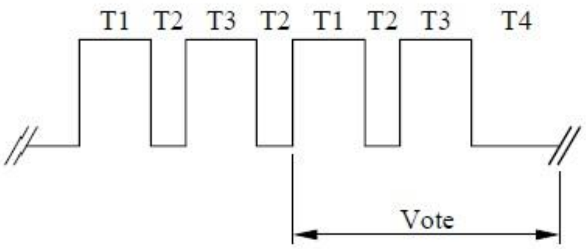
The laboratory environment should reflect the home environment in terms of light condition, visual threshold and reflections. The test material have to correspond to the assessment problem, were a wide range of material address overall performance investigations and a specific range of material address a certain weakness of the system. It should at least attend non-expert 15 observers in a subjective test. They must be tested to clarify if the subject has normal eye vision, conceiving depth information and colour vision. Preliminary findings suggest that non-expert observers may yield more critical results with exposure to higher quality transmission and display technologies.
Subjects should be carefully introduced to the test method, quality factors that are likely to occur, the grading scale, sequence and timing. The test material should be at a random order for normalization.

2.4.1 STIMULUS

Single, double and comparison stimulus are the main types of stimulus. There are many subtypes of these were information are presented differently. These subtypes are further explained in ITU-R BT. 500-11 [32].

The single stimulus (SS) method uses only one initialization reference before the whole test itself while the double uses a reference between each test picture.

The double-stimulus (EBU) method is cyclic in that the assessor is first presented with an unimpaired reference, then with the same picture impaired. Following this, the subject is asked to vote on the second, keeping in mind the first.



b) Variant II

Phases of presentation:

T1 = 10 s Reference picture

T2 = 3 s Mid-grey produced by a video level of around 200 mV

T3 = 10 s Test condition

T4 = 5-11 s Mid-grey

Experience suggests that extending the periods T1 and T3 beyond 10 s does not improve the assessors' ability to grade the pictures or sequences.

Figure 7. In the Stimulus-comparison method, two images or sequences of images are displayed and the viewer provides an index of the relation between the two presentations [32].

2.5 PEARSON CORRELATION

The Pearson product-moment correlation coefficient is a measure of linear dependence between two variables X and Y. The correlation coefficient is defined as the covariance of the two variables divided the product of their standard deviations. Function 2 defines the correlation coefficient between the variables X and Y [15]. This correlation method was applied in section 6.5

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \qquad (2)$$

3: SUBJECTIVE TEST SETUP

The subjective experiment was performed at NTNU September/October 2011 in the 3D laboratory at the department of Electronics and Telecommunications in collaboration with Q2S. The setup of the subjective experiment is one of the main objectives in this report. Since subjective measurements for auto-stereoscopic are not yet standardized, this is a very important aspect of the test. This section will derive different setups and technology that was tested to reach the final setup.

3.1 AUTO-STEREOSCOPIC DISPLAY

A 'Tridelity MV4200' was used in the experiment [31]. This is an auto-stereoscopic display utilizes the parallax barriers technology. The screen is 42" and has a native resolution of 1936×1360 pixels. It supports 5 and 9 views. In this project it was used 5 views, because it is less complex and the results may be easier to interpret. The view pattern goes like ⋯ V3, V4, V5, V1, V2, V3, V4, V5, V1, V2, V3 ⋯ horizontally distributed. This view pattern from V1 to V5 are repeated 5 to 6 times at 3.592m from the screen along the horizontally axis. The distance between each view will vary according to the distance to the screen. It was measured that the interval between the neighbour views is 47mm, 63mm, and 72mm at the viewing distance of 2.595m, 3.592m, and 4.587m, respectively. This can be observed in the left image in figure 8.

In the transition area between V5 and V1the observer will perceive the most amount of crosstalk, since one eye sees these two views at the same time. This is because V1 and V5 have the largest disparity between the captured images. The optimal viewing distance is 3.5m with a view interval of about 62mm [31]. Both view interval and system crosstalk pattern from neighbour views vary along the viewing distance. Figure 8 illustrates the captured images by a camera when view V3 was totally white and the other 4 views were totally black. It can be observed that the white illumination is not perfect, since some dark areas are visible. It is easy to see that the system crosstalk is more dominate at close distances, crosstalk are black colour area in, a). It becomes better at the optimal distance, b) and increases at longer distances again. It is typical that the observer will perceive more crosstalk in positions closer than 3.5m compared to further away. The objective capture of the screen is in detailed explained in section 5.

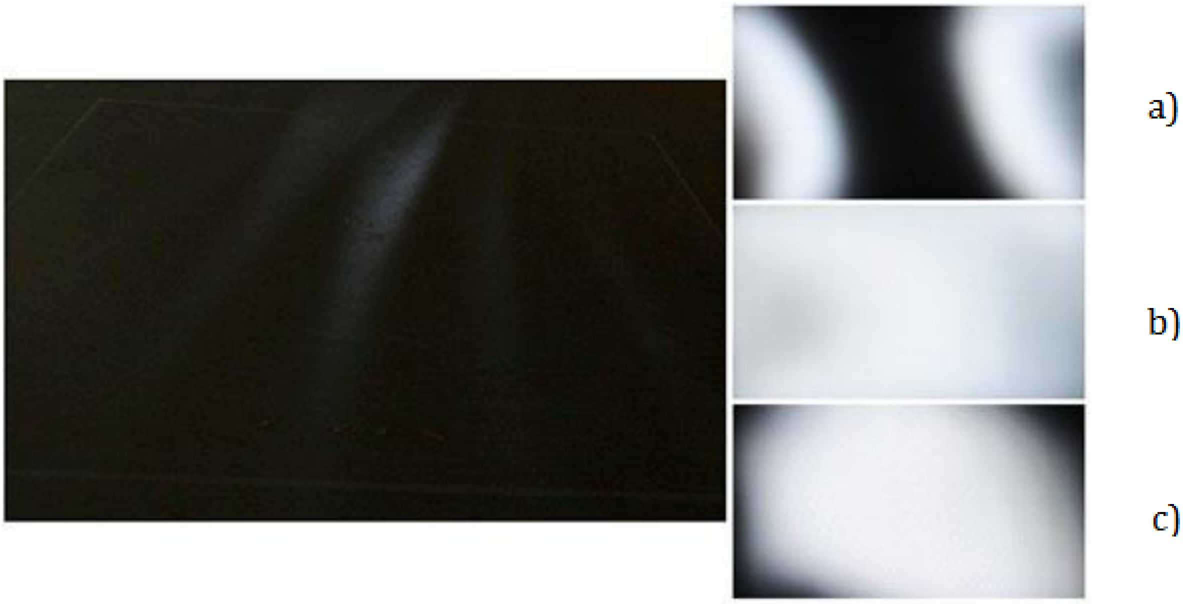


Figure 8. Left image: Capture of the moving area with only V3 radiance. The light radiance are wider at a close range compared to further away a) System crosstalk at 2.595m b) System crosstalk at 3.592m c) System crosstalk at 4.587m.

The measured colour temperature, luminance and gamut of the screen were 6500k, 95.5cd/m^2 and 2.2. This was executed with an EyeOne Display2 sensor. The input format has to be similar to figure 9 in order to display 5 views.

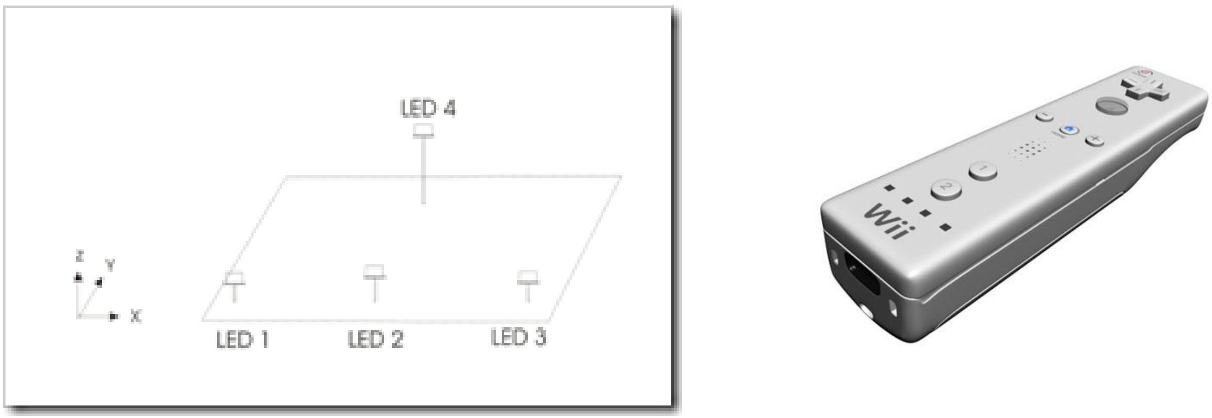
Figure 9. Input format of 5 views.

3.2 HEAD TRACKING SYSTEM

In order to detect the subject during movement in the selected movement area a tracking system was created. The goal was to extract the centre position between the eyes of the subject during the test. It is quite hard to get a perfect position of the subject with the available tracking technology, especially without spending a lot of money on the tracking system. There are many different ways of getting this position with different technologies. Two different approaches were tested to obtain the subjects position. Both approaches are based on open source software. This is an advantage since it makes it easier to adjust the tracking system.

3.2.1 WIIMOTE VIRTUAL REALITY DESKTOP (VR)

The Wiimote has a resolution of 1,024 × 768 pixels, more than 4 bits of dot size or light intensity, a 100 Hz refresh rate, and a 45 degree horizontal field of view. The core of the VR setup is using the Nintendo Wiimote controller as a tracking camera, illustrated in figure 10 b). The Wiimote has an infra-red camera that is able to recognise up to four infra-red lights. The idea was to create a device with four infra-red LEDs which can be detected by the Wiimote controller. The integrated algorithm reconstructs the position of the created device [21]. The infra-red LEDs on the device have to be mounted as figure 10 a) shows. This is necessary for obtaining the entire three axes that reveals the subjects position. In order to get the position in to our computer it was necessary to have a Bluetooth receiver and rewrite the Wiimote software to print the position of the IR device. This rewrite process in C is detailed in [21], and will not be further explained.



a) b)

Figure 10. a) Displays the orientation of the IR lights [21] b) Wiimote controller [30]

Figure 11. The IR sensor system attached to a helmet.

3.2.2 KINECT PRIME SENSOR™ NITE 1.3 FRAMEWORK

The Kinect system consists of two depth sensors and one RGB camera which is lined horizontally on the base structure. The sensors and the camera operate at a frame rate of 30Hz and 8-bit VGA resolution of 640×480 pixels. The software enables advanced gesture recognition, facial recognition and tracking. The sensor boundaries go roughly from 0.8m to 6m in the depth direction. The vertical and horizontal fields of the Kinect sensor are about 63cm and 87cm at the distance of 0.8m. This results in a resolution of 1.3mm per pixel. The angular boundaries are 57° horizontally and 43° vertically [24]. The horizontal accuracy at 5m was 3cm deviation when the subject moved 1m to ether of sides. The sensor uses a steady detection system to indentify the subject and a skeleton tracking system to keep track of the subject. The basic assumption of skeleton tracking is that the user's upper body is generally inside the field of view. The tracking algorithm uses the label map output by a user segmentation process to generate a skeleton. In the indentify subject part, the skeleton model adjusts and calibrates the tracking frame with the subject. The output of the skeleton system is the position and orientation of the skeleton joints. Only the head position is of interest in this project [28]. The position is subtracted out as x, y and z coordinates, were x is the horizontal alignment, y is the height of the subject and z is the distance to the subject. The software was adapted to integrate with a numpad, implied the coordinates are stored for each time a number from 1 to 5 are pressed. The user of the system gets discarded after 10 seconds, if the subject leaves the detection area [28].

Figure 12. The Kinect Prime Sensor™ system[26]

3.3 SCORE TRACKING

The subjects were supposed to give a score from numerous random positions during the test. In order to do so the subject used a Logitech diNovo Media Desktop Bluetooth Numpad to enter the scores. It is wireless and got buttons for from 0 to 9. When a score are typed by the subject, both the score and the head position from the head tracking system was recorded into a file.

3.4 SCENE CONTENT

The perception of crosstalk is influenced by a wide range of depth structures [18], image contrast [11, 14, 18], edges [11, 18], textures [11, 18] and pixel disparity. According to these different picture elements, nine different pictures form MPEG [19] were selected based on different image variations and disparity between the views. The pictures used for training was selected in such a way that each quality level was represented. The whole range of quality levels within the set of test stimuli was covered. The largest disparity level between neighbouring views was 67, 12, 8 pixels for the training pictures and 8, 18, 22, 30, 38, 40 pixels for the test pictures. These first three pictures in figure 13 where used to training before the test and the bottom six pictures in figure 13 were used during the test.

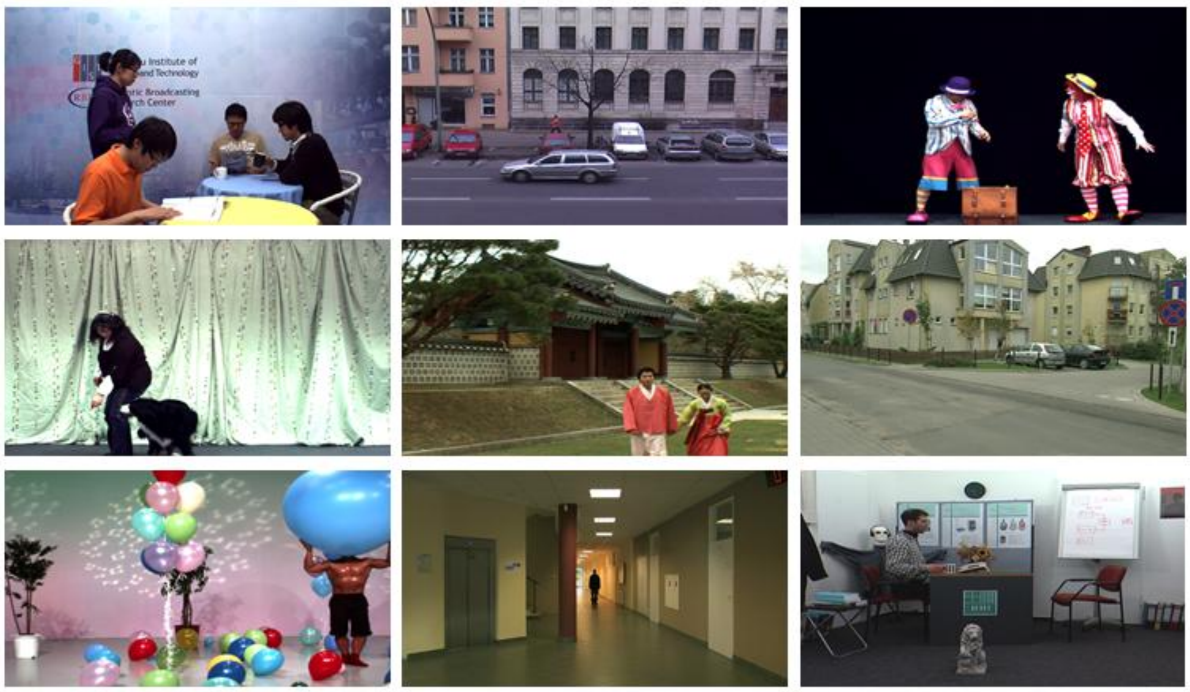


Figure 13. The selected scenes used in the project. From top left to bottom right: Cafe, Outdoor, Pantomime, Dog, Love Bird, Poznan Street, Balloons, Poznan Hall, Book Arrival. Among which, the top row is for training.

3.5 LAB ENVIRONMENT

The allowed moving region for the subjects in front of the auto-stereoscopic display was restricted in a 2.0m×2.0m square region. The main reasons for this restriction are angle limitations of the Kinect tracking system and the optimal viewing distance of the auto-stereoscopic display. If the subject moves too long outside the allowed area, the tracking system looses connection with the subject. If this situation occurred during the test, the subject had to repeat the identify user process in order to continue the test. This process was explained in section 3.2.2.

The orientation of the square region is described as followed; The Z axis is the depth. This means the distance from the display to the subject head position. The X coordinate describes the subjects horizontally head location movements and the Y coordinates describes the vertically head location movements. The moving region has X values from -1.0m to 1.0m and Z values from 2.5m to 4.5m. Summarized the XZ plane are horizontally parallel with the marked square region on the floor, while the Y axis varies with the subjects height alteration while sitting during the test. The Y position was fixed to 126.5cm above the floor, the same level as the display centre during the test. The goal was to have the same fixed eye position, 126.5cm above the floor, for every subject. This was done by adjusting the height of the chair in relation to the subject personal height. Five same-height anchor points was used to calibrate the subject height before the test. Four anchor points marks the four corners of the moving region and the last one was placed in the middle of the back corners, as illustrated in figure 14. The subject had to move to all these positions, sitting in the chair, use the Bluetooth numpad to type the zero score button three times fast in a row. This was used to make an average XYZ position for each of the five anchors. These values created a virtual identification plane for each subject. If anything was moved during or between tests like for example the Kinect sensor was accidentally moved 1.0cm in some direction, we could use the virtual plane data to shift the values to the right ones.

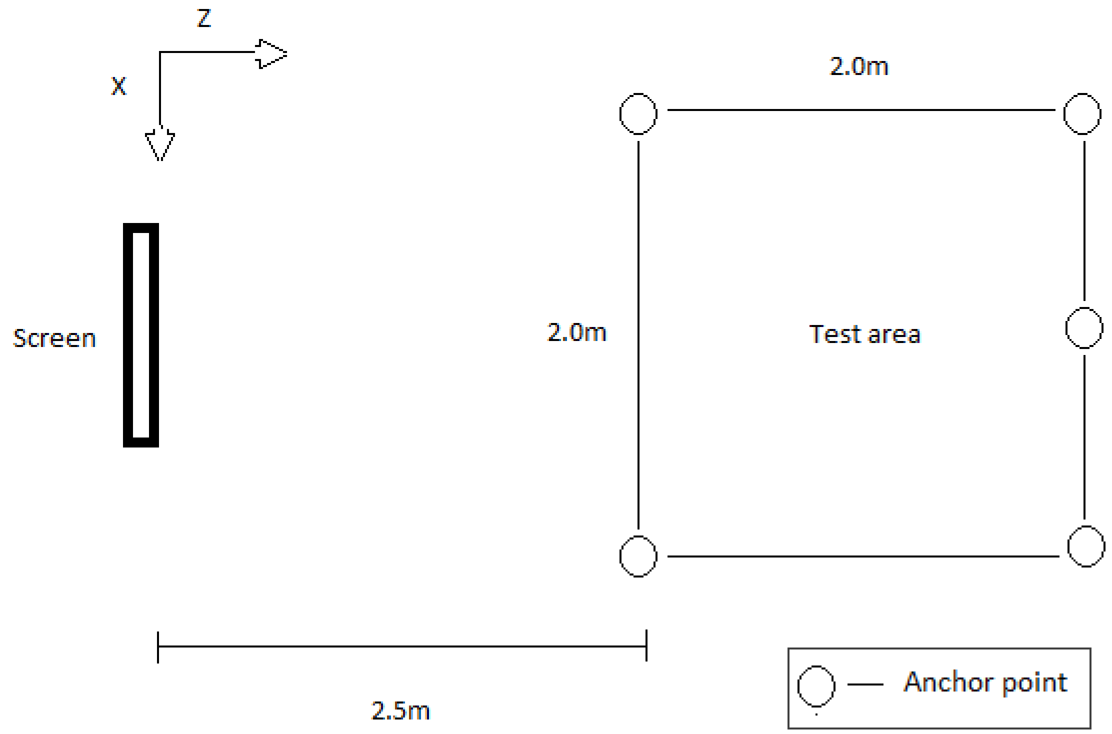


Figure 14. Directions and test area seen from above**.**

The sitting chair has five wheels. This makes it quite easy to move freely inside the restricted area while sitting in the chair. Regardless of the static Y position it will differ a little since a human being cannot sit totally still during movement**.**

All windows, door and other small light sources were covered by black curtains during the test. This is for avoiding light reflections from other light sources. It was only used two reading lamps as background light in the test room. We used the EyeOne Display2 to measure colour temperature and illuminance of the background which were respectively 2500k and 63 lux. Figure 15 shows the test setup.

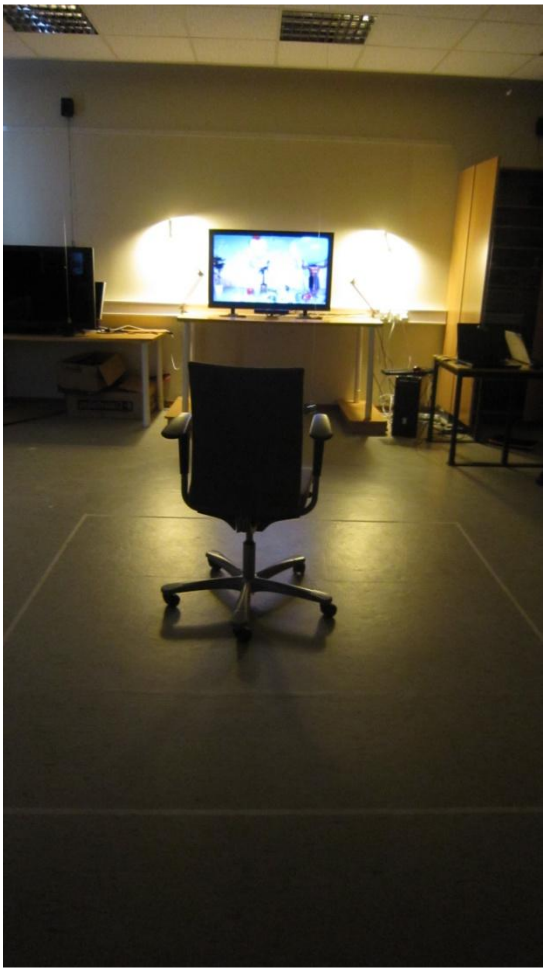


Figure 15. The test setup. This picture shows the marked area; the chair which was uses and distance relation. The background light is much brighter in this picture compared to reality. The reason for this is the camera settings that were used.

4: SUBJECTIVE CROSSTALK MESUREMENT

This section will describe the test methodology that was used to conduct the subjective crosstalk assessment. It will describe the stimulus, subjects, training and test session.

4.1 SUBJECTS AND VISUAL SCREENING

25 test subjects participated in this test. The gender distribution was 15 males and 10 females. The age varied from 21 to 50 years old, with a average of 27.56 years and standard deviation of 7.38. Of 25 subjects did 9 of them use eye correction aids. The subjects were from Europe and Asia.

All the subjects had to complete a visual screening. The screening checked the colour vision, the ability to see stereoscopic depth and normal eye vision. All the data which was collected are completely anonymous. Every subject got an id number.

4.2 SINGLE STIMULUS

Since there is no existence of reference pictures for crosstalk perception, a double stimulus method cannot be adopted. The subjects will always perceived system crosstalk by the auto-stereoscopic display. The single stimulus was also used for assessing the quality levels of stereoscopic images with varying camera baselines in the literature [16], [18] when is it hard to decide the reference.

4.3 TRAINING SESSIONS

Each subject had to complete a training session to get an idea of how to evaluate the test stimuli. Every subject got a fast introduction of what crosstalk is and just an overview to understand why they doing the test. The subjects were told to notice the amount of crosstalk for each practice picture. They were guided to the best and worst possible positions for each training pictures. The mean opinion score table (MOS) from 1 to 5 was used to score the test pictures. The score of 5 is the highest score and 1 is the lowest, this explained in table 1.

The subjects were first shown the training picture named Cafe, which in the worst positions had a score 1 and at the best possible position a score 2. The subject could move freely in the surveyed area for 90 seconds to get an overall impression of the score 1 ‘very annoying’ and 2 ‘Annoying’. During these 90 seconds the test leader explained how to observe and evaluate it. The test subjects got really fast an idea of what to look for. In the Cafe picture, the test subjects said it was easy to see the ghosting around certain objects like the table, coffee cups and around the persons in the picture. By ghosting we mean that certain content in the picture has a partly transparent shadow of itself, as shown in figure 4. The ghosting can often appear at one side, but sometimes at both sides of the viewed object. In that case you'll see three representations of the same object. Some subjects said it was rather annoying to stare at this picture for a longer period, and that they really had to concentrate to see the content clearly.

The same was repeated for training picture 2 named Outdoor, which had a score 3 'slightly annoying' at the worst position and a score 4 'perceptible but not annoying' at the best position. This picture was shown for 60 seconds. The last training picture named Pantomime has a score 5 'imperceptible' at the best position and about a score 3 in the worst. This picture was shown for 45 seconds.

The training pictures were looped for 15 seconds each after the first round, until the subject had a good understanding of the scores from 1 to 5 and how to distinguish the scores from each other. The test subject was told to ignore colour changes and other abnormalities like the moiré effect that were not related to crosstalk. The training session continued until the subject could understand and distinguish between the five different quality levels.

	Explanation	Example
1	Very annoying: Crosstalk is so much, that the 3D perception is hardly to be formed and you feel a little uncomfortable.	Café - Worst
2	Annoying: Crosstalk is much, although the 3D perception still can be formed but you refuse to accept viewing such quality in daily life.	Café - Best
3	Slightly annoying: there is obvious crosstalk. However you can accept viewing such quality reluctantly.	Outdoor - Worst
4	Perceptible but not annoying: you can see a little bit crosstalk at a first glance, but the quality of the whole image is still good.	Outdoor - Best
5	Imperceptible: you cannot see any crosstalk or you can perceive very slightly only when you pay special attention to a certain region.	Pantomime - Best

Table 1. Explanations of five categorical adjectival levels and their training examples

4.4 TEST SESSIONS

The test session was divided into six sub sessions. Each subject got to evaluate one picture for each sub session, in total six pictures. Each picture was displayed for 7 minutes. The pictures were randomly chosen from the test stimuli (Dog, Love Bird, Poznan Street, Balloons, Poznan Hall and Book Arrival). The subject was positioned in a chair which was height-adjusted for every subject. They were told to move randomly in the surveyed area and evaluate the picture with the MOS scale based on a first impression. The subjects applied the training session to make a score for each position. The test subject was encouraged to randomly move around without following any static patterns or searching for high or low values patterns. It was important to try to judge the quality without consider it too much. Just trust their first feeling. It was also important that the subject tried as best as possible to sit in the same sitting-position before entering a score, since there is many ways to sit in a chair. Each subject tried to cover as many positions as possible.

5: CREATING CROSSTALK-DISTORTED PICTURES

This section will describe several steps that were required in order to create a distorted image based on objective capture of crosstalk. The images that were created with this method were used as an input value to find an objective measure of crosstalk. The crosstalk pictures were compared to its original picture with a structural similarity algorithm. This structural similarity algorithm (SSIM) was described in the section 2.3. The method described below is used to create all the objective images that were used in this project.

5.1 CREATING THE PICTURE SETUP

In order to capture crosstalk it was necessary to find a method that reveals crosstalk added to neighbouring views. It is desirable to detect the amount of the unwanted neighbouring views that appears in the current view. The basic idea was to create a layout where only one of the views was totally white (rgb:255) and the rest of the views where black (rgb:0). When this setup was displayed on the auto-stereoscopic display the white leakage from the single white view is easy detectable on the rest of the views. Figure 16 shows the input format that was given the auto-stereoscopic display. The expectation of this example was that white pixels from V3 would be detectable in specially V2 and V4 since they are the closest neighbouring views. Some leakage will be detectable to V1 and V5 as well. Investigating white leakage into a black picture are much more efficient and more accurate than if this method had used the actual colour images to find the crosstalk. This is because measuring crosstalk in captured images of the screen introduces several sources of error from camera parameters that influence the colour, the projection itself and errors occurring from the utilized comparison algorithm.

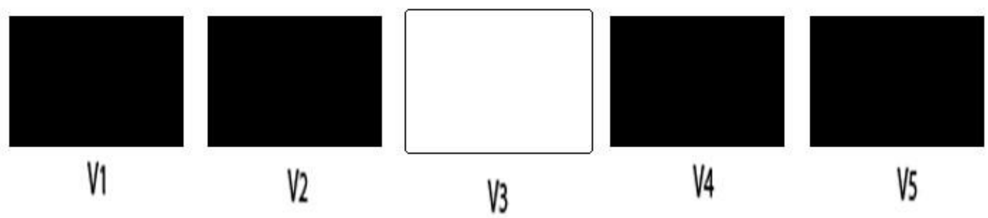


Figure 16. Picture setup: a) V1 b) V2 c) V3 d) V4 e) V5

Pictures were captured by a camera to intercept the crosstalk. The camera was shifted to a best as possible position that revealed V3 as white as possible. The rest of the 4 views were captured related to V3 in order to intercept crosstalk. The pictures were taken at 2.595m, 3,592m and 4.587m from the display, in the z direction. The distance between each picture (the x distance), were 47mm, 63mm and 72mm corresponding to given distances. The described sequence was repeated for every view. Figure 17 show examples of the captured images. In the example were V3 totally white and the rest black. ***The amount of crosstalk is depended of the transparency of the black and white pixels*.**

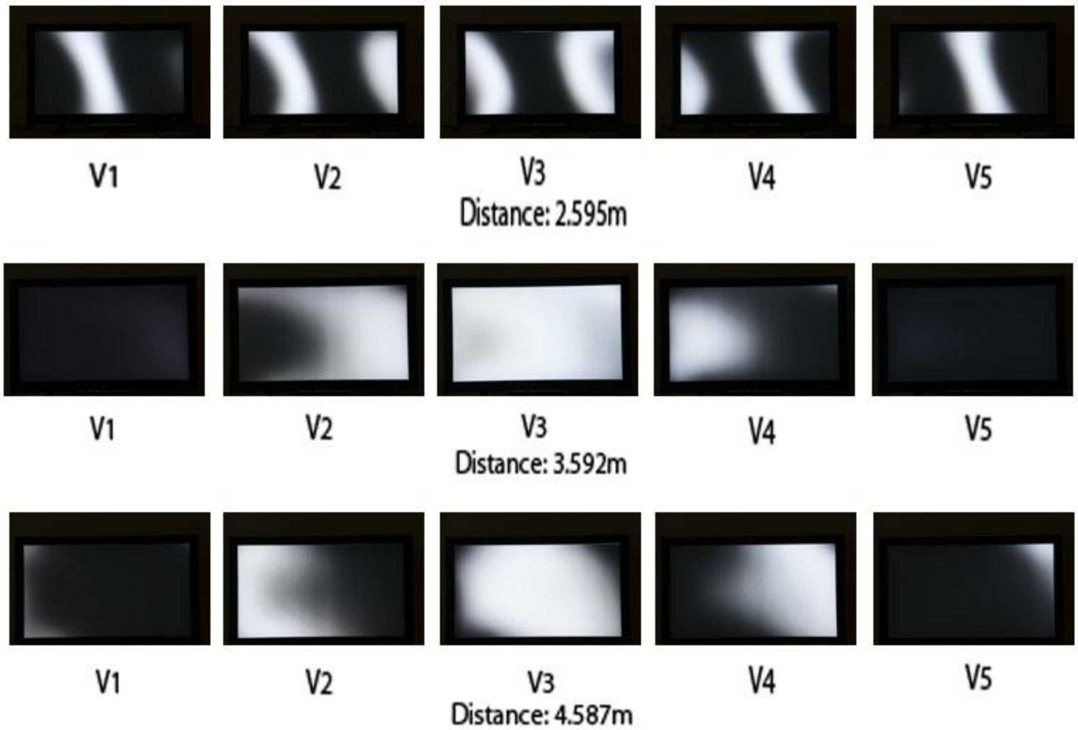


Figure 17. Captures from different distances

The optimal result will be that V3 will be totally white and the rest of the views totally black. It is easy to observe that V3 at 3.592m are nearly optimal, but there are some dark shadows in the picture as well. V2 and V4 prove a lot of crosstalk emitted from V3. It can be observed that V3 at 4.587 have more elements of black in the corners, but still quite good. The crosstalk distribution at V2 and V4 at 4.587 is different compared to V2 and V4 at 3.592m. The crosstalk distribution at 2.595m is very hard to catch. The emitted light prisms at this distance are so narrow and it results in a lot of crosstalk, especially in terms of perceived crosstalk for the observer. The observer will perceive more than two views at a close distance which results in a perception with a lot of disturbance.

Figure 18 show the distribution of the captured pictured when the reference picture (the totally white picture) is altered from V1 to V5 at 3.592m. The first observation is that crosstalk is nearly equally disturbed when the white reference picture are altered. This can be observed by comparing Row1 – V2, Row2 – V3, Row3 – V4, Row4 – V5 and Row5 – V1 in figure 18 below.

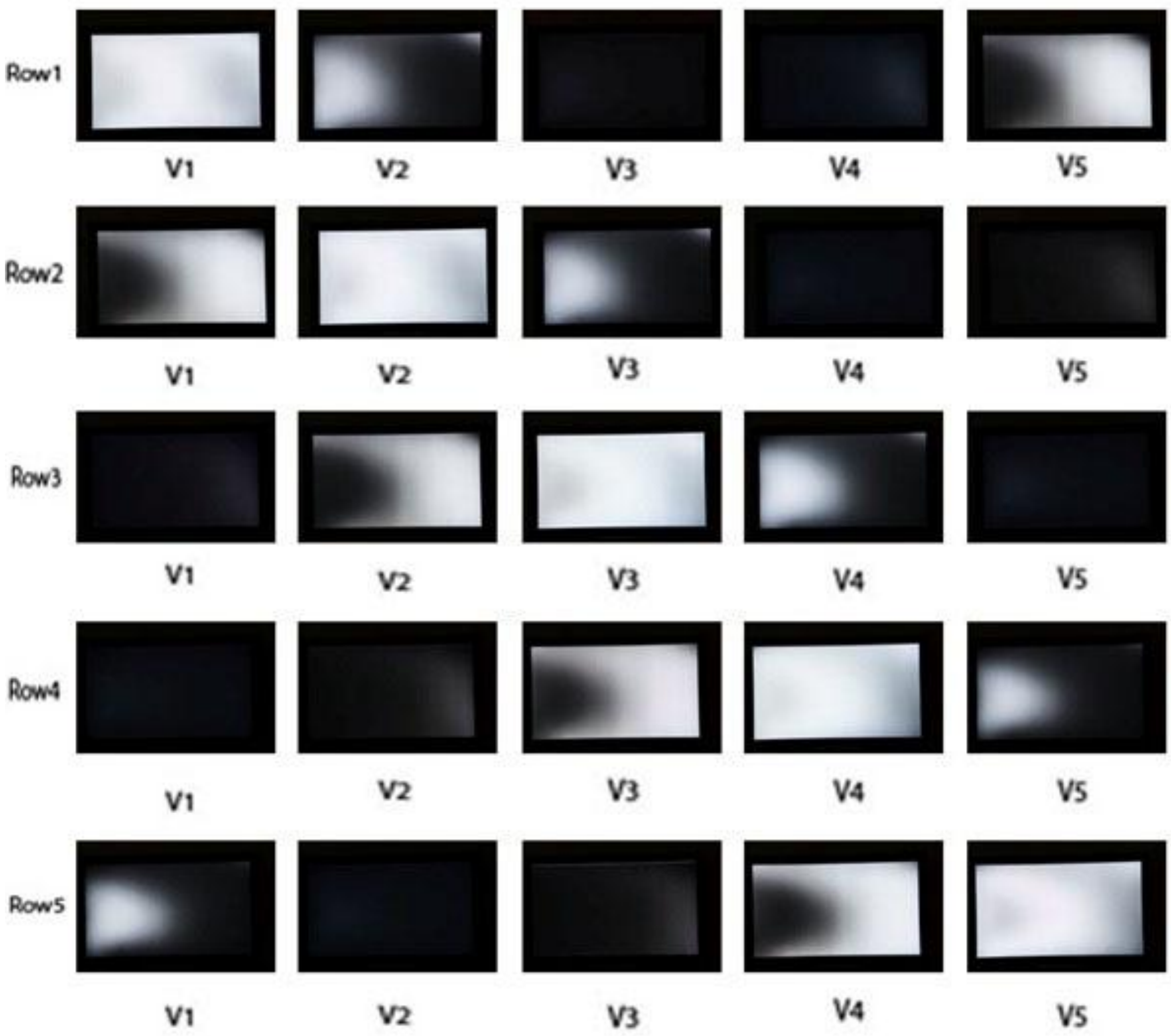


Figure 18. Distribution of the captured pictured when the reference picture are altered from V1 to V5

The reason that this method functions is that crosstalk is equally produced for a specific display for any input format. The crosstalk distribution will only deviate from different displays and viewing angles. It is important to differ between the static crosstalk distribution given from a display and the amount of perceived crosstalk for an observer. As we know are the perceived crosstalk influenced by a number of elements, as described in section 2.2, 3.1 and 3.4. While the derived method to not suffer from the content in a picture.

These captured images generate a base for creating pictures with crosstalk. The amount, intensity and position of the crosstalk are now revealed. This information can be used to create crosstalk influenced pictures that consist of two pictures. Crosstalk is more complex than adding unwanted parts from a neighbouring view to the current one***. Crosstalk in a single view is not fully determined by leakage from one neighbouring view, but a combination of all the views. To simplify the creation of objective crosstalk, the crosstalk in this project was determined by the neighbour that influenced the current view the most.***

The creation of an objective crosstalk influenced picture includes several steps. The first step is to remove the outer unwanted edges of the picture. The information of the picture is only what is emitted from the screen itself. In addition was a black background added, since the refined image is a little tilted to the left. The implemented SSIM algorithm compares only images with the same resolution. The refined pictures are then scaled down from 4046x2251 pixels to; 1024x768 pixels for the Balloons, Book arrival and Lovebird pictures, 1280x960 for the Dog picture and 1920x1088 for the Porznan Street and Porznan Hall pictures. The reason for the down-scaling was that the refined pictures resolution had to be equal to the native resolution of the original images. It is generally better to down-scale the least complex picture instead of up-scaling a complex image, since up-scaling invents new pixels while downscaling just removes pixels.

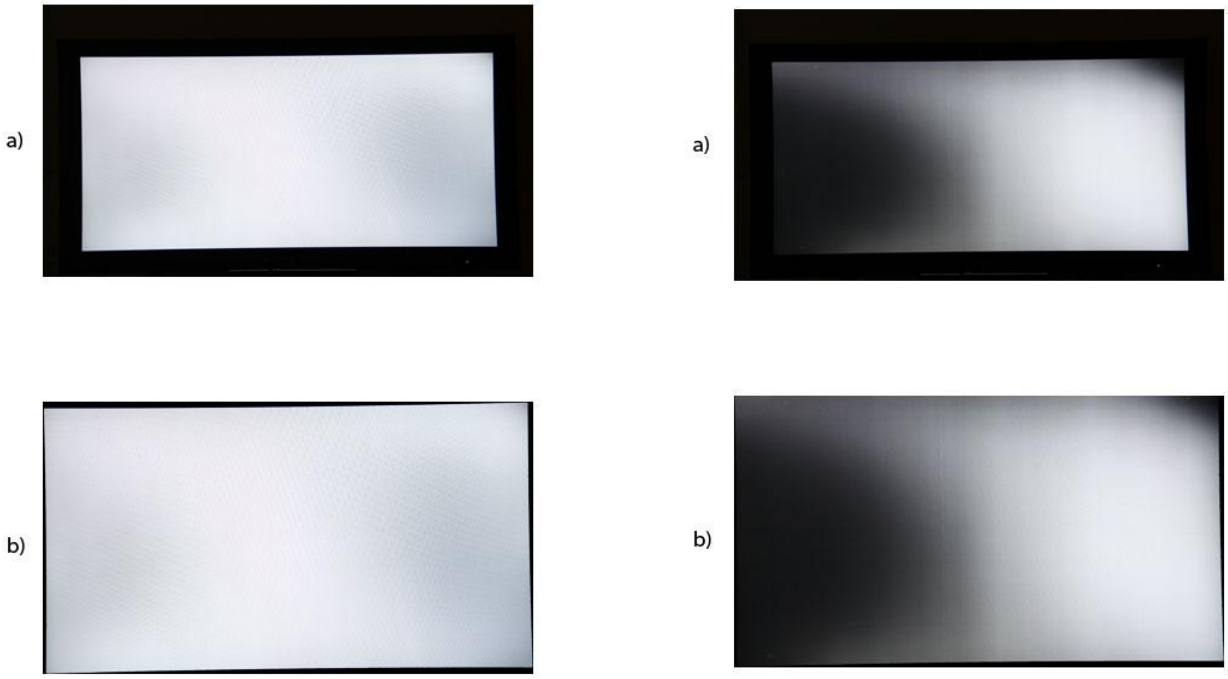


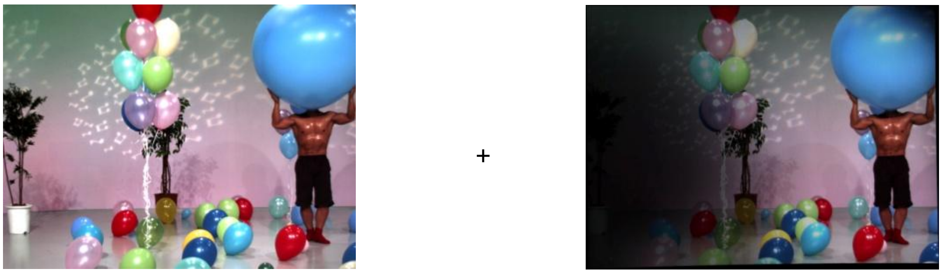
Figure 19. a) Original capture. b) Refined edges

The white area in V4 and V2 in Figure 19 indicates the influence of V3 in V4 and V2. This white area is now a basis for a mask of V3 which will later be added into V2 or V4. The next step now is to apply this mask to the original image that represents V3. Figure 20 gives an example of the masked picture. The example is taken from the "balloons" picture. The image in Figure 20 c) is the amount of the unintended V3 that influence V2 in the example.

a) V2 b) original V3 c)

Figure 20. Process creating objective pictures

The dark area of Figure 20 a) and c) indicates where the original image of V2 is. The image from Figure 20 c) is added to the original image of V2 to complete the stereoscopic image, where crosstalk from V3 is added into V2. The colour intensity indicates how transparent the crosstalk is.



V2 V3

Figure 21. V2 and V3 are added together

Figure 22 is an example of the final objective crosstalk created picture. It is easy to observe crosstalk at the right side of the picture, which correspond to the theory this method for creating crosstalk was based on. Figure 23 shows a SSIM map where the crosstalk applies. An important observation is that crosstalk is easier detectable around edges of objects in the picture compared to homogenous areas of the picture. Crosstalk is also more prominent on objects in the foreground than objects in the background.

Figure22. Example of an objective created crosstalk influenced picture.

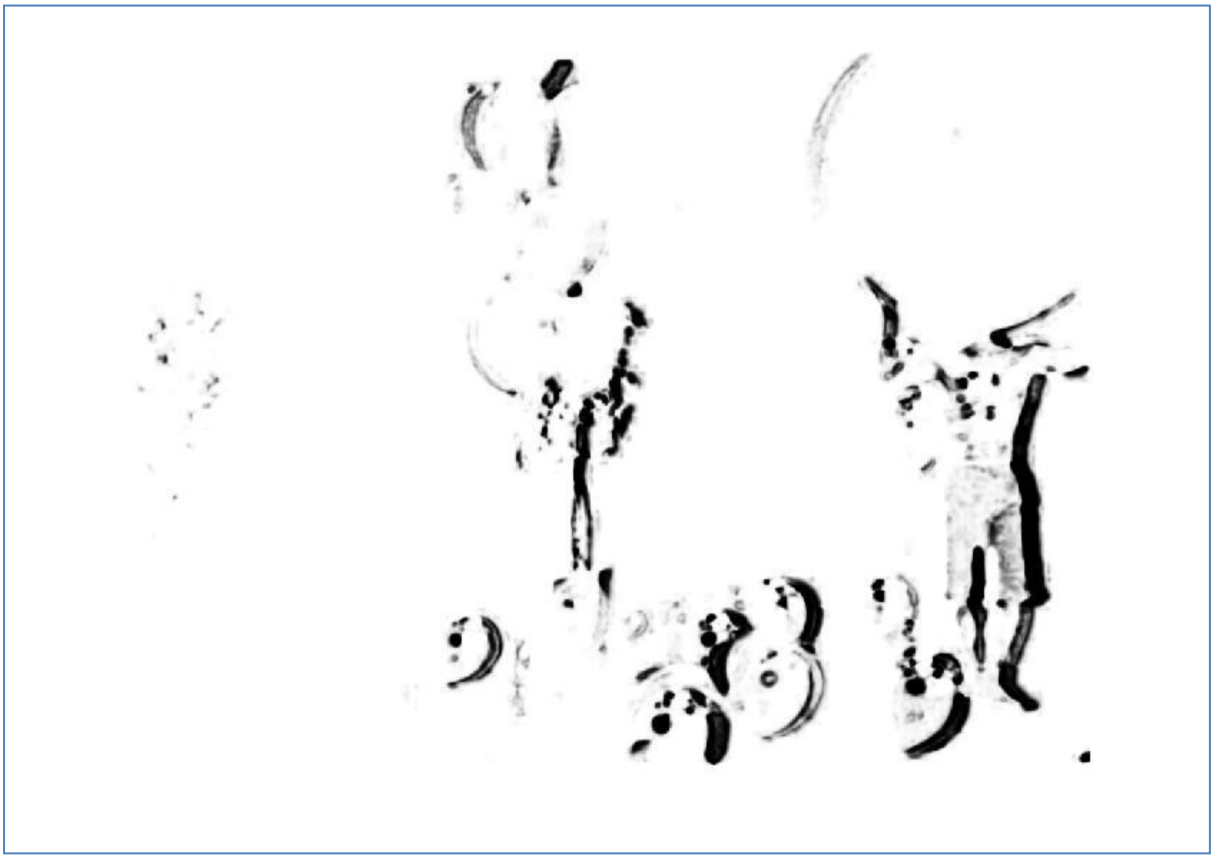


Figure 23. SSIM map shows where crosstalk applies.

5.2 APPLYING THE SSIM ALGORITHM – RESULTS

The SSIM algorithm was created in Mat.lab. It was based on the implementation that was created by Matthew Gaubatz [29] and modified in order to work with my current scripts, indexing and variables.

The SSIM algorithm needs two input parameters, a reference image and a distorted image. The structural similarity index are calculated based the input parameters and leaves a single SSIM value for the whole distorted image. This value represents the objective measure of crosstalk in a picture. This value is related to the position, since the crosstalk captured images have a position attached to it. The basic theory about SSIM was derived in section 2.3

5.3 EXPECTATION OF THE CROSSTALK DISTRIBUTION

In section 3.1 second paragraph, it is mentioned that the observer will perceive the largest degree of crosstalk in the transition area between V1 and V5. This area is within about 6.3cm at 3.592m distance from the screen. The reason for this is that the disparity from V1 to V5 is the largest. This transition area is greater in a 5 view picture with a large disparity degree between each view compared for a picture with a smaller disparity degree. The comparison of the disparity degree between the pictures Dog and Book Arrival are examples on small and large disparity degree.

The crosstalk difference between V1/V2, V2/V3, V3/V4, and V4/V5 are expected to be nearly the same. The exception is if there are some content that is greatly influenced by crosstalk that only appears in only a fraction of the views. An example of this is a foreground object that only appears in a fraction of views. The stop sign at the right side in Poznan Street image may have a bigger influence between V4 and V5 than V5 and V1. The main reason for this is that the black and white crosstalk distribution in section 5.1 shows that in this case crosstalk in a picture is more influenced by the right side neighbour than the left one at the distance of 3.592m. This may only apply for the Tridelity MV4200' screen.

I have derived a simple hypothesis that creates a model based on the expectation deduced in the paragraph above. Figure 24 show my model that illustrates that the crosstalk influence is greater between V1 and V5 (Distance: L) than for V1/V2, V2/V3, V3/V4, V4/V5 with distance (4L) at 3.592m from the screen, in the Z direction. Figure 24 can be seen as a perfect slice at 3.592m in the X direction.

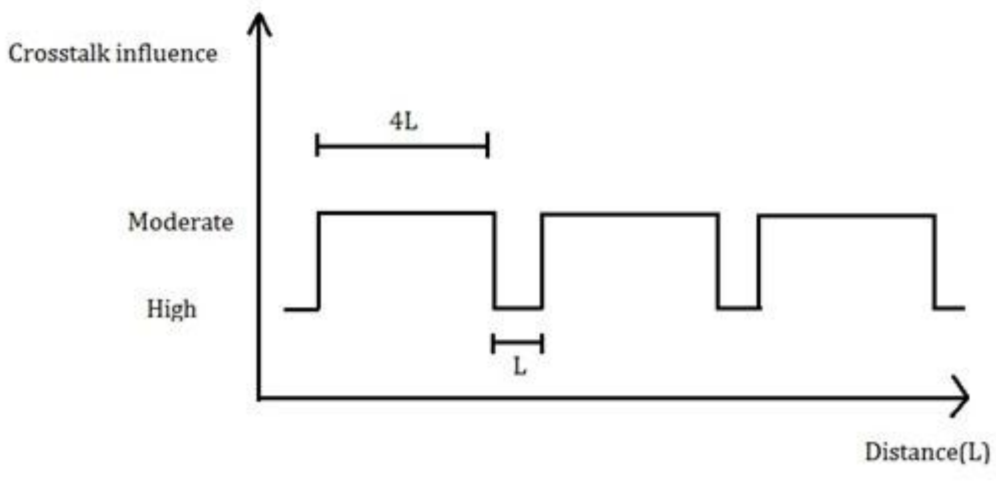
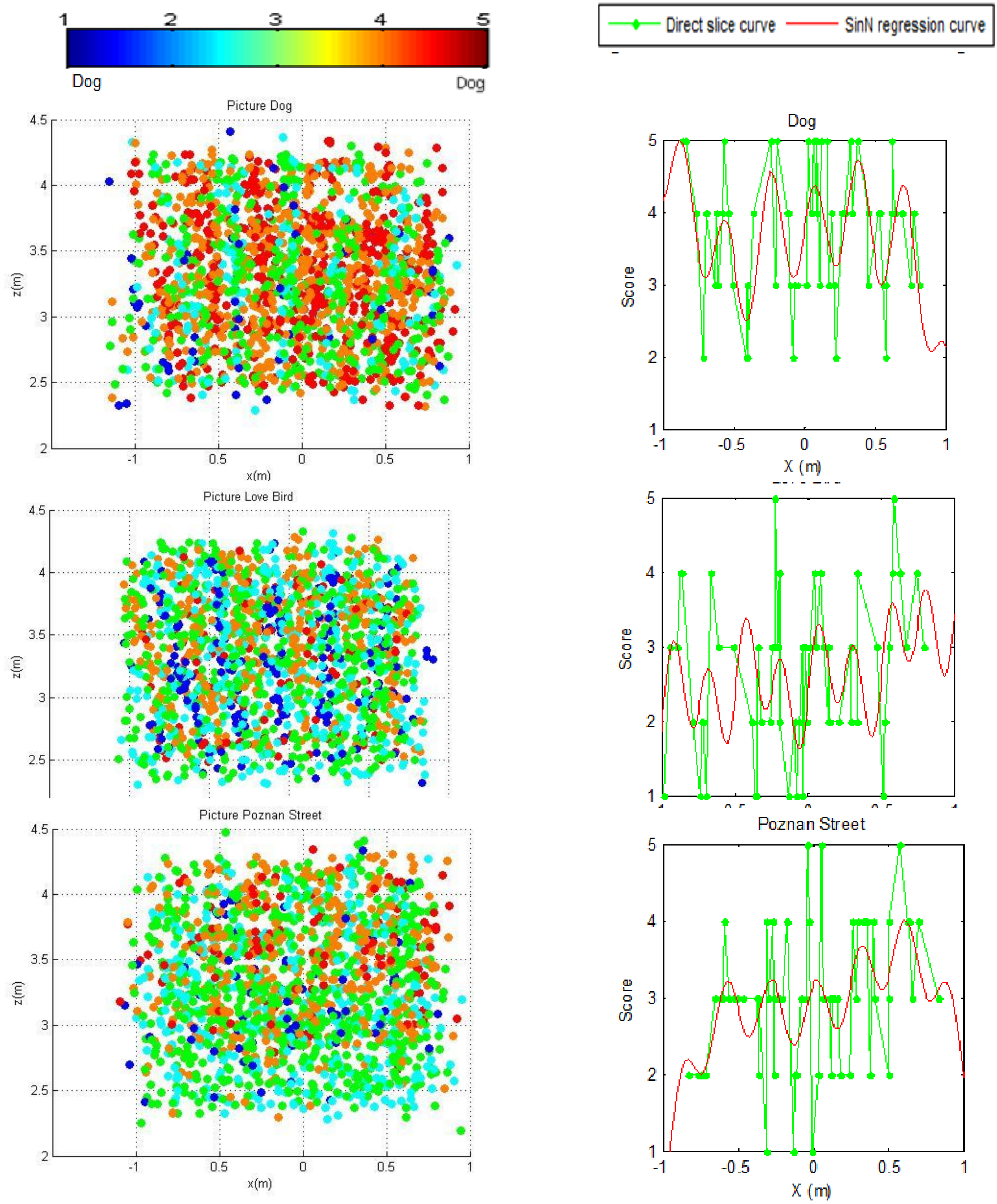


Figure 24. Model of hypothesis derived above

6: RESULTS OF SUBJECTIVE TEST

6.1 RAW DATA ANALYZE

The raw data approach is based only on the raw subjective scores obtained from the subjects. These data are not interpolated in any way. Here is the basic thought to use the mind to find patterns of the data that correspond to the objective expectations. The plots below shows the raw data plot to the left and a slice of the plot at 3.5m to the right, which is at the optimal viewing distance. A sine regression curve was created to get a better visual idea of the pattern.



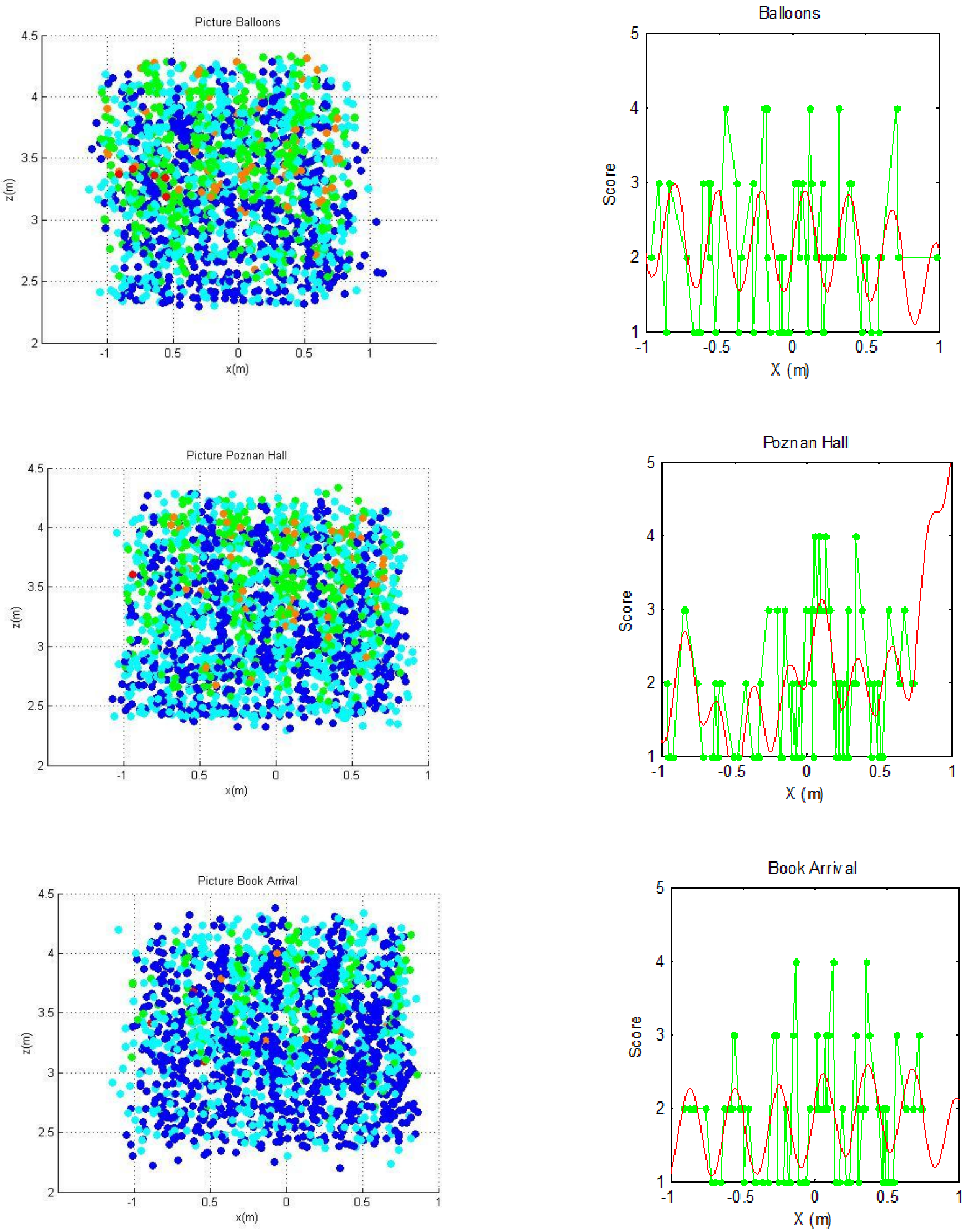
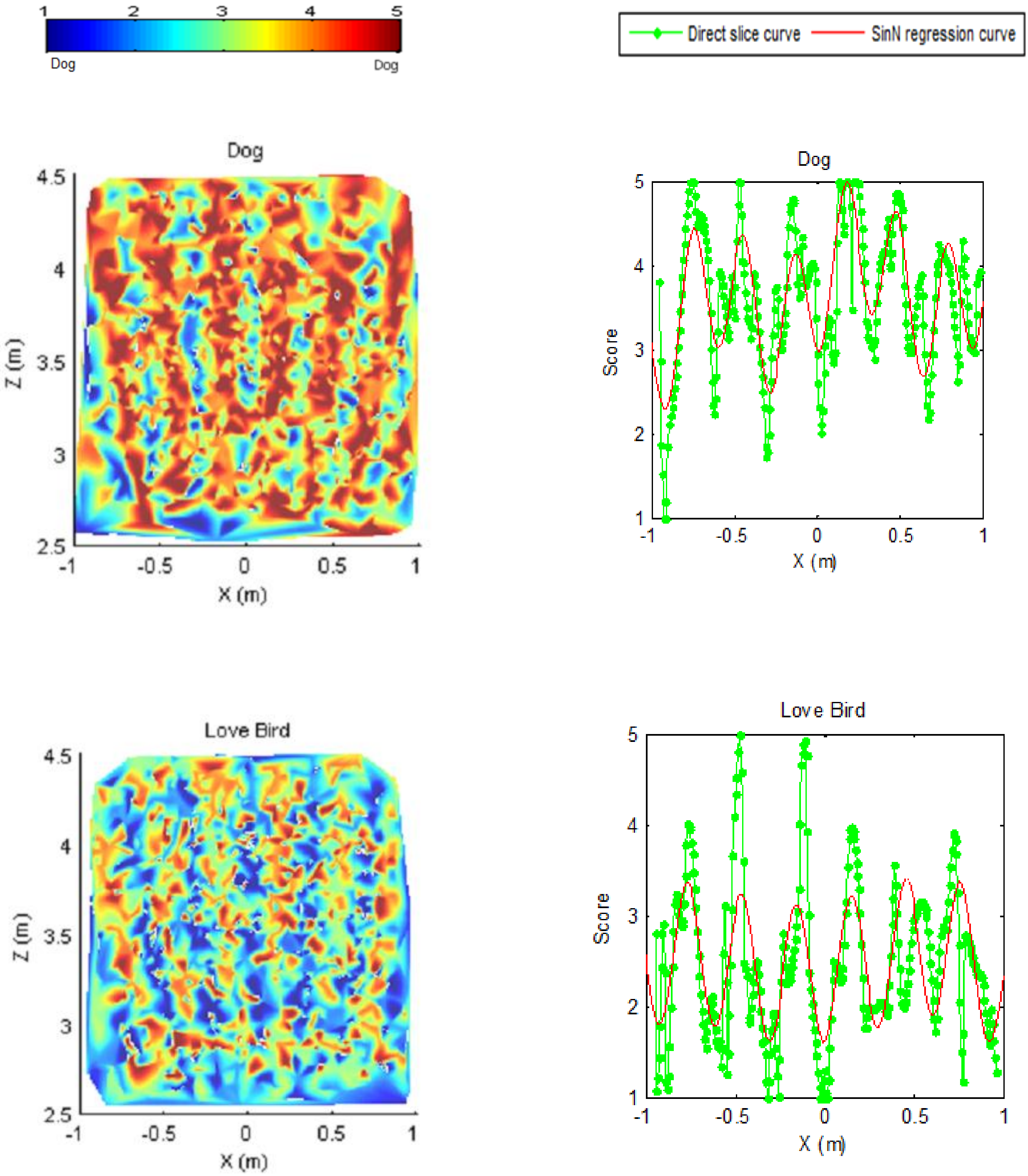
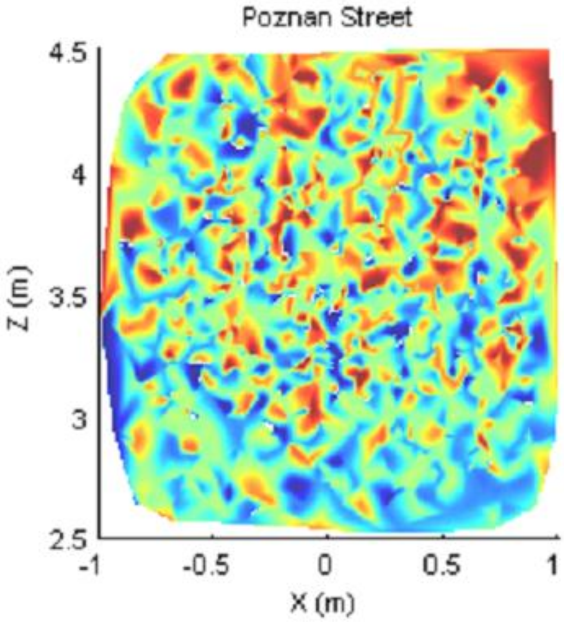


Figure 25. Raw data results.

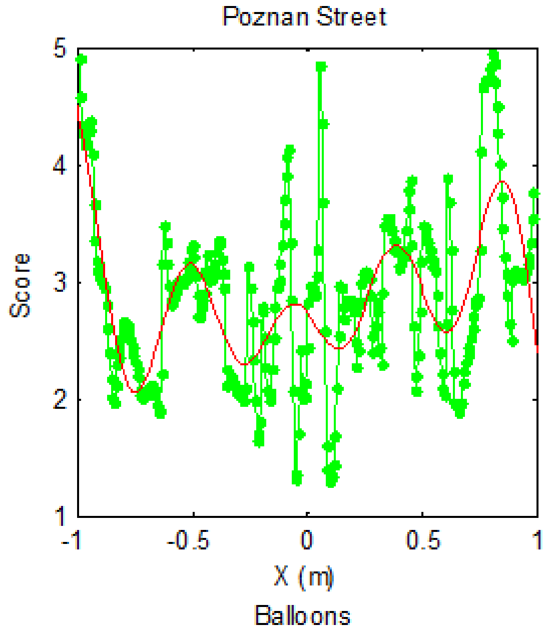
6.2 INTERPOLATED DATA ANALYZE

The interpolated data is based on the raw data to create a surface. The interpolation uses the raw data to generate fictitious data points between the raw data points to make a surface. There are several different interpolation methods like linear, nearest neighbour and cubic. The cubic method offers true continuity between the raw data scores. It is based on the cubic equation $y = Ax^3 + Bx^2 + Cx + D$ where the volume of an object is proportional to Ax^3, object's surface area is proportional to Bx^2 and the Cx is proportional with the size [23].

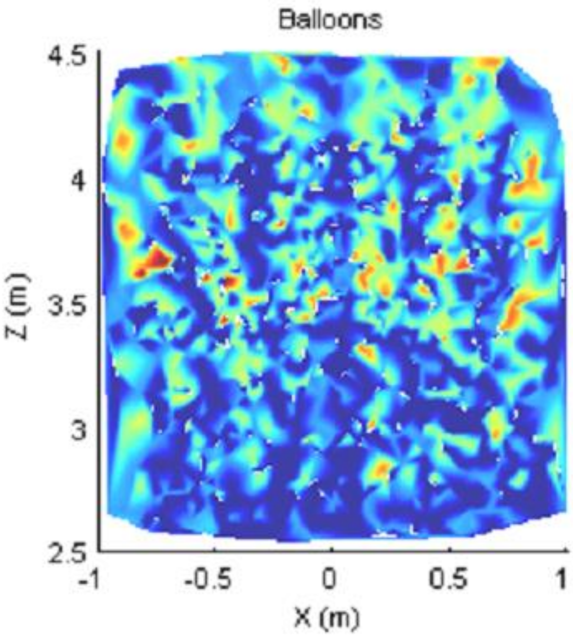




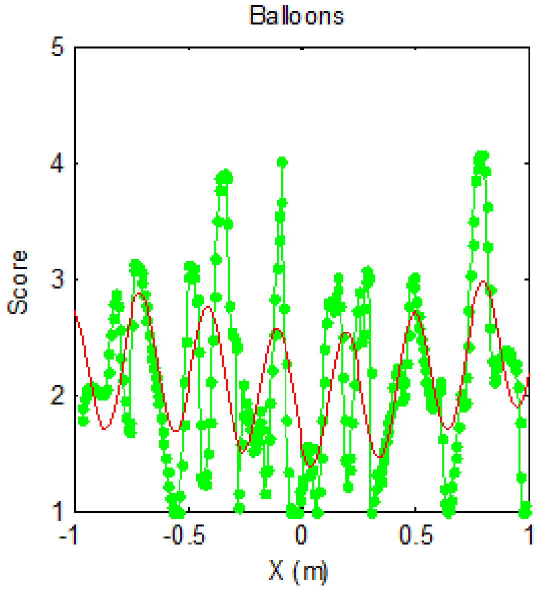
Poznan Street
Z (m)
X (m)
4.5
4
3.5
3
2.5
-1
-0.5
0
0.5
1



Poznan Street
Score
X (m)
Balloons
5
4
3
2
1
-1
-0.5
0
0.5
1



Balloons
Z (m)
X (m)
4.5
4
3.5
3
2.5
-1
-0.5
0
0.5
1



Balloons
Score
X (m)
5
4
3
2
1
-1
-0.5
0
0.5
1

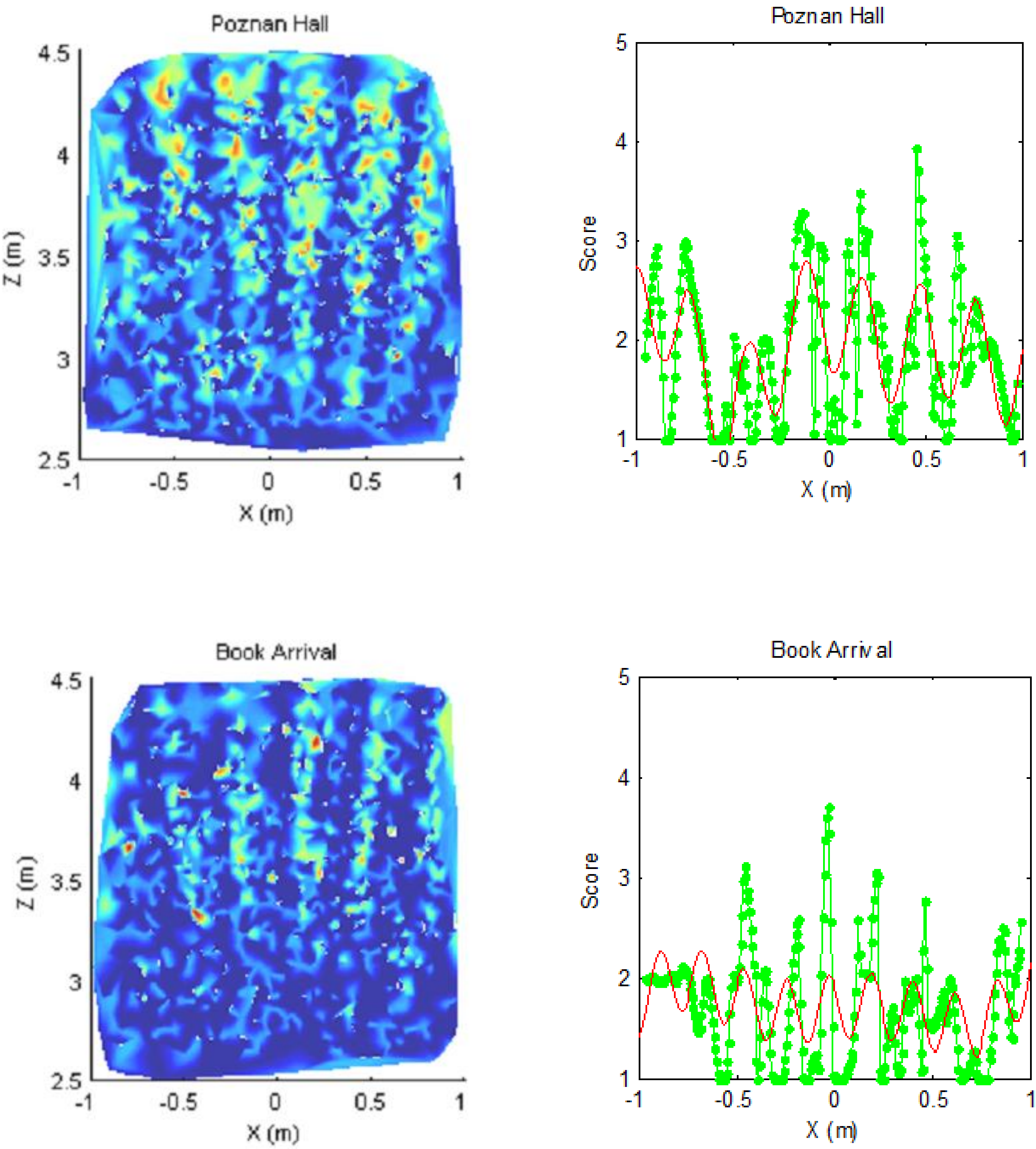


Figure 26. Interpolated data results

The table 2 below shows basic statistics for the score distribution for each picture in the subjective test. The intension of this table is discussed in 7.2 .These values are extracted from appendix 3.

Picture	Dog	Lovebird	Poznan Street	Balloons	Poznan Hall	Book Arrival
Mean	3,7041	2,8364	2,9455	1,9604	1,8711	1,5767
Std.Dev	1,0744	1,1163	1,0156	0,8944	0,8107	0,7208

Table 2. Means and standard deviation of the values given by subjects

6.3 EXTRACTION OF SUBJECTIVE VALUES

This section will explain how subjective values are extracted from the subjective test results. These subjective values will be compared with objective SSIM values to observe the correlation. The extraction of subjective values is one of the main concerns in this experiment. The way the subjective values are extracted will substantially affect the correlation. It is proposed two different methods for extraction values in this section, and it supports the issues around extraction of subjective values.

The research question is; *"How to extract subjective values from a specific position that will with a high probability correlate with the objective values created by the SSIM algorithm?"*

For simplification reasons are the distance of 3.592m (+/- 5cm) chosen to be investigated for the rest of the experiment. The crosstalk influence is much more complex at a close distance and more ambiguous at a further distance.

The main problem is that the subjective values are finite and randomly spread in an area of roughly 2x2 meters. It is essentially that subjective values can be distinguished by a minimum 6 cm, since the distance between two views at 3.592 meters are around 6.3 cm. This minimum distance is even smaller for distances closer to the screen. There are many positions in the raw subjective data that are not covered by this requirement. This problem is the reason for that the subjective results had to be interpolated in order to extract subjective values with a precision better than 6cm. The model derived in section 5.3 with figure 24 illustrates the perfect scenario that the subjective results should appear like at the distance 3.592m.

6.4 INTERPOLATION

The first interpolation is based on a sine regression curve. The sine regression curve is fitted to mimic the periodic pattern in figure 24. A five order sine regression curve (function 3) was applied to the subjective data, were a, b and c are coefficients (15 in total for each wave) of the sine wave.

$$Y(x) = a_1 \sin(b_1 x + c_1) + \ldots.. + a_5 \sin(b_5 x + c_5) \qquad (3)$$

The figures in appendix 1 show the resulting sine waves for each picture at 3.592m.

One period for each sine wave is selected to represent the subjective values that are compared with the objective SSIM values. The period was selected based on the highest density of the subjective values. The period is marked with a green rectangle in appendix 1. This sine period are sampled into 5 values, where the first value are set at the zero point of the curve. The next four values are sampled with a distance of 6.3cm with decreasing x values. This distance is the same distance of the captured objective pictures taken of the screen. Each sample gives a score value that represent the subjective value in a certain position. An example of the sampling is given in figure 27. Their selected periods are all shown in appendix 1, where the pink marks represent a sample.

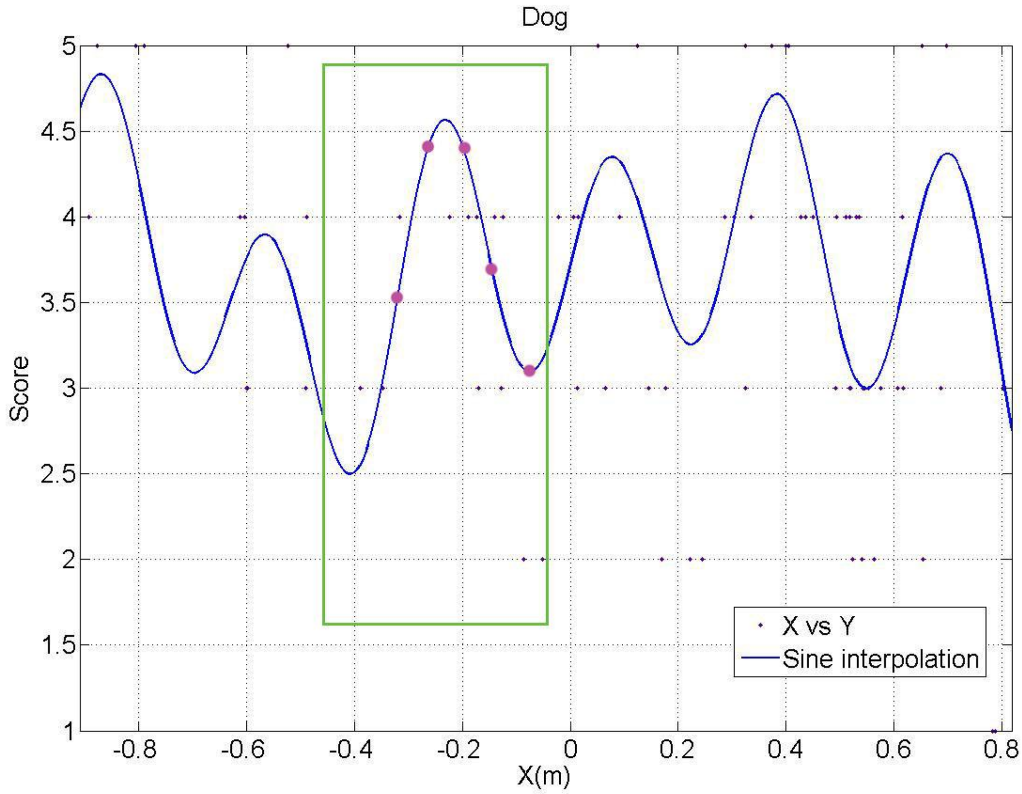


Figure 27. Sine wave with sampling

The second extraction method uses both a sine interpolation and a linear interpolation to model the hypothesis derived in section 5.3. The method expand the area where the crosstalk influence is moderate compared to the periodic sine wave in the first method. It is desirable to approach the moderate crosstalk influenced area to a distance of 4L compared to the highly crosstalk influenced area with distance L. In the first method, moderate and highly crosstalk influenced areas had an extent with equal distance. The consequence is that sampled values near the starting value have too low score, according to the derived model in section 5.3.

In the second method we try to equalize this by taking a linear interpolation into account. The effect of this is that the sampling area contain more high score values around the top value of the sine graph. The specific extraction rule are as follow: The sine wave itself gives initially the value, but if a linear interpolated value are higher than the sine value in a given position, the linear interpolated value are used. The exception is if the linear interpolated value are higher than the maximum sine value (the value of the top point) for the entire sine period, the linear interpolated value will be discarded and replaced with the sine value in the given point. Figure 28 below describes the sampling visually. Note the samples attached to the linear interpolation which apply to the rules derived above.

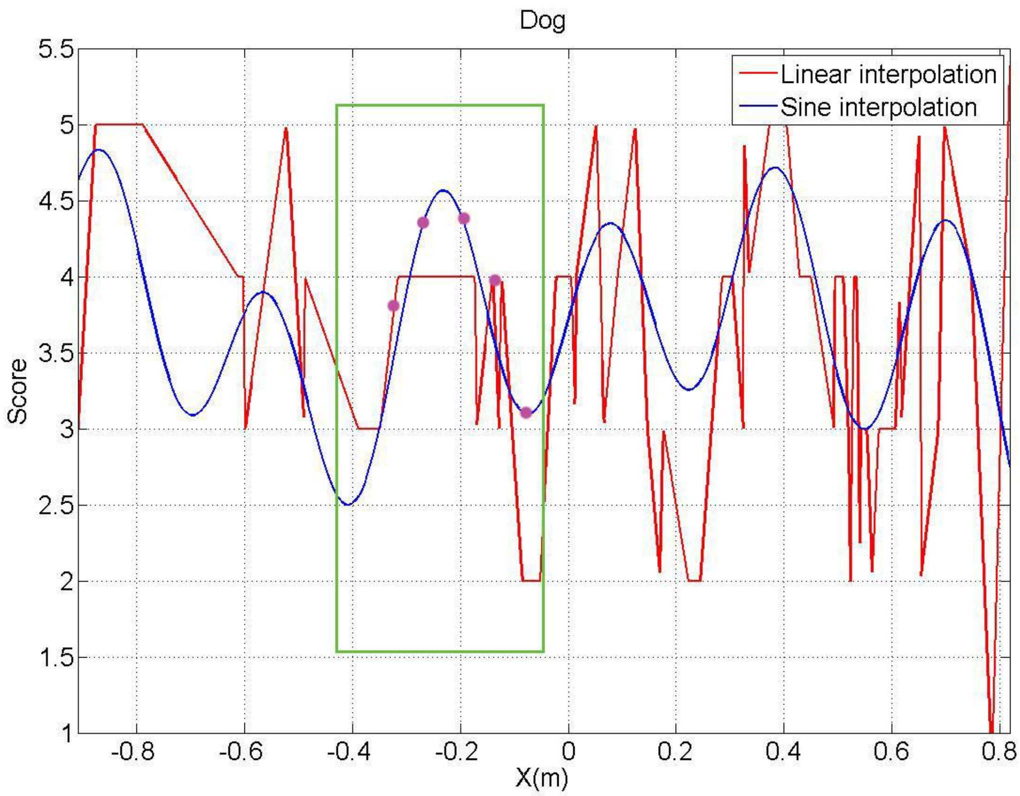
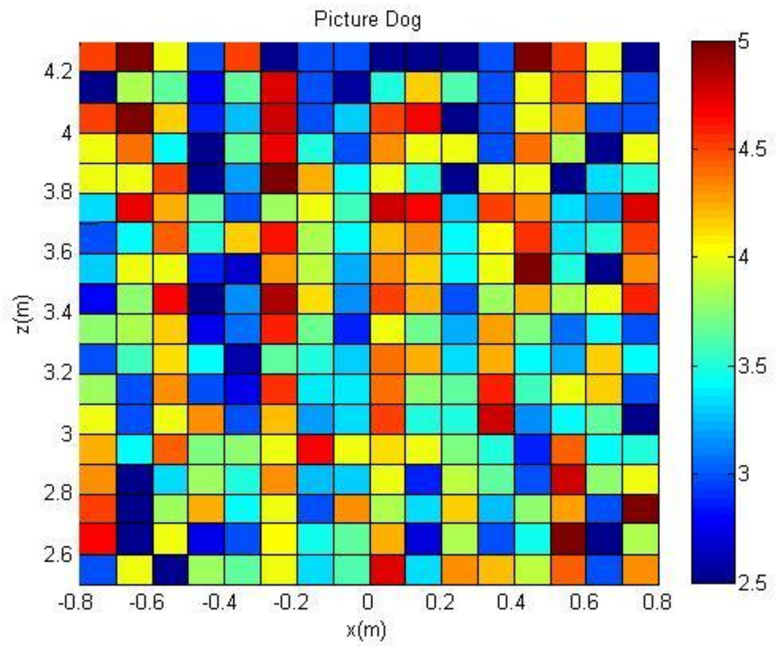


Figure 28. Sine wave and linear interpolation with sampling.

6.4.1 TILE PLOT (ALTERNATIVE APPROACH)

An alternative approach of interpolation was to create a map of tiles were each tile represented a mean score of an area of 10cm. The figure below shows the result of the tiles were the colours are heat related. Red and yellow indicate a score mean between 4 and 5 while blue is around a mean score of 3. The problem with this approach was that the density of user scores was too low, especially around the edges. It is still possible to see the periodic pattern marked in white in figure 29 b). This method got discarded because the lack of user scores gave a bad correlation. This method can work with a higher density of user scores which implies it is possible to make smaller tiles < 5cm.

a)



b)

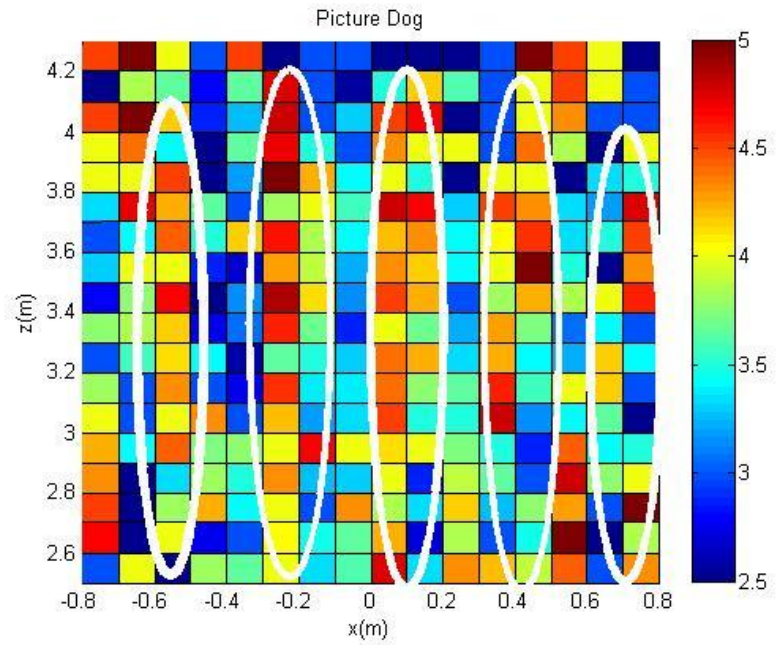


Figure 29. a) Tile plot of subjective scores, were each tile represent the mean. b) shows the periods.

6.5 RESULTS OF OBJECTIVE MEASUREMENT AND IT'S CORRELATION WITH SUBJECTIVE RESULTS

The section is divided into two parts. The first is the result of the objective measurements. It shows the SSIM values and PSNR values for each picture and for all the pictures together. The second part provides the results of the correlation analysis of the two extraction methods derived in section 6.3 above.

6.5.1 OBJECTIVE MEASUREMENTS OF THE OBJECTIVE CROSSTALK CREATED IMAGES.

Table 3 shows the objective values that are used to find the correlation with the subjective values.

Picture: Dog	SSIM	PSNR	Picture:Balloons	SSIM	PSNR
V1InfbyV2	0,998053	45,007315	V1InfbyV2	0,974146	29,908937
V2InfbyV3	0,999197	50,114504	V2InfbyV3	0,976123	30,216110
V3InfbyV4	0,996749	43,562477	V3InfbyV4	0,976910	30,847858
V4InfbyV5	0,999198	48,434773	V4InfbyV5	0,975065	30,115925
V5InfbyV1	0,994519	40,921738	V5InfbyV1	0,951327	23,147987
Picture: Lovebird			Picture: Poznan Hall		
V1InfbyV2	0,983466	34,970483	V1InfbyV2	0,991548	35,510400
V2InfbyV3	0,989053	37,066995	V2InfbyV3	0,990441	35,935299
V3InfbyV4	0,986158	34,924536	V3InfbyV4	0,992894	36,048763
V4InfbyV5	0,987512	36,282436	V4InfbyV5	0,992903	36,320088
V5InfbyV1	0,970418	31,800865	V5InfbyV1	0,978815	29,494103
Picture:PoznanStreet			Picture: Book Arrival		
V1InfbyV2	0,963176	30,369919	V1InfbyV2	0,991548	35,510400
V2InfbyV3	0,986183	35,046212	V2InfbyV3	0,990441	35,935299
V3InfbyV4	0,989445	36,388219	V3InfbyV4	0,992894	36,048763
V4InfbyV5	0,987896	35,510664	V4InfbyV5	0,992903	36,320088
V5InfbyV1	0,973298	31,864087	V5InfbyV1	0,978815	29,494103

Table 3. The SSIM and PSNR values extracted from objective crosstalk pictures

There are two important results to note. The first is that the SSIM and the PSNR value in which V5 are influenced by V1 gives the worst result for all the pictures except for picture Poznan Street. The second observation is that the results of V1infbyV2-...-V4infbyV5 are quite similar, except for picture Poznan Street. These results will be discussed in section 7.7.

6.5.2 RESULTS OF CORRELATION ANALYSIS

The first correlation analysis is the correlation between the objective SSIM/PSNR values and the interpolated sine curve given in appendix 1.

Picture	SSIM Correlation (Pearson)	PSNR Correlation (Pearson)
Dog	0,4631	0,3600
Lovebird	0,7926	0,6440
Poznan Street	0,6558	0,7441
Balloons	0,3571	0,3723
Poznan Hall	0,2585	0,2610
Book Arrival	0,4026	0,3622
All pictures	0,4288	0,7086

Table 4. Correlation based on only the interpolated sine wave

The second correlation analysis is the correlation between the objective SSIM/PSNR values and the interpolated sine curve and linear interpolation together to mimic the model in section 5.3. The graphs are given in appendix 2.

Picture	SSIM Correlation (Pearson)	PSNR Correlation (Pearson)
Dog	0,7380	0,6958
Lovebird	0,8774	0,7420
Poznan Street	0,5913	0,6855
Balloons	0,7416	0,7549
Poznan Hall	0,5780	0,5191
Book Arrival	0,6812	0,7178
All pictures	0,4955	0,6709

Table 5. Correlation based on the sine wave and the linear interpolation

Since the SSIM values are not linear, the SSIM values are parameterized with function 4 below to get a better correlation.

$$Y(x) = \frac{a}{1+e^{(-b(x-c))}} \quad (4)$$

If a, b and c are select to be 1, 2 and 3 the resulting correlation changes.

Picture	SSIM Corr. - model	SSIM P-value model
Dog	0,8981	0,0385
Lovebird	0,9561	0,0110
Poznan Street	0,6310	0,2536
Balloons	0,7608	0,0135
Poznan Hall	0,5946	0,2903
Book Arrival	0,7157	0,1740
All pictures	0,5139	0,0037

Table 6. Correlation based on the sine wave and the linear interpolation with parameterized SSIM values

It is important to observe that the correlation improves compared to the results in table 6. If the two pictures Poznan Hall and Book Arrival with the highest disparity are excluded from the experiment, that is Dog, Lovebird, Poznan Hall and Balloons still remains, gives an improved result. The combined result gives a correlation of 0,8249. Figure 30 shows a plot of the correlation.

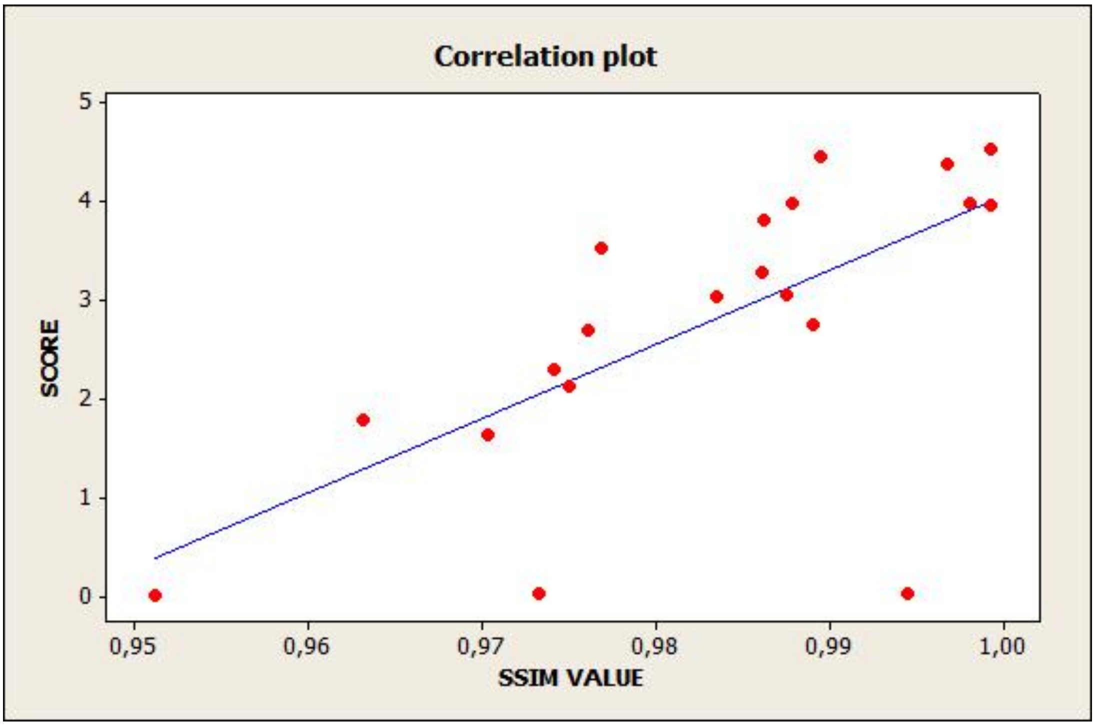


Figure 30. Correlation plot of all the pictures except from Poznan Hall and Book Arrival

7: DISCUSSION

The discussion section is divided into two parts; Subjective experiment and objective analysis.

The discussion will analyse the method, materials and results that was made or used for this project. It will go through the setup and execution of the test step by step and analyze the outcome of the results and suggest improvements.

Subjective part

7.1 GOAL, RESTRICTIONS AND EXPECTATIONS OF SUBJECTIVE TEST

The goal was to map the characterisation of the auto-stereoscopic screen based on subjective data. A perfect map will have infinite number of score data related to the position. This is one a clear restriction of subjective data. A subjective result will always suffer from inadequate data at some positions. That is why interpolation is an option for creating data.

All the subjects had their own random moving path in the viewing region. The result of this was that the subjects did not cover exactly same position compared to each other. The concept of mean opinion score (MOS), confidential interval (CI) and screening of subjects defined in the ITU-R BT.500 [20], which is a common method for result analysis, could not be applied in this case.

The expectation was that the subjective results would be very fuzzy, almost impossible to interpret. There were many clear sources of error that could affect the result, which will be discussed in the next sections of the report. It was a little surprising that the results turned out pretty good for this first attempt. There are many actions that can improve the result as well. Some of them will be further explained.

7.2 ANALYSIS OF SUBJECTIVE RESULTS

The scores are related to the viewing position and follow certain distribution according to characteristics of the auto-stereoscopic display. It can be seen from both raw data and interpolated plots that subjective data follows a pattern periodically related to the viewing position. This periodically patterns are quite clear along the X axis. These patterns are easiest to observe from the raw data plots and its slice plot in figure 25 and 26 for the Dog, Balloons, Poznan Hall and Book Arrival. The periodically score pattern of the pictures Dog and Love Bird are pretty clearly in the interpolated plots with its corresponding slice plot. The patterns of the image Poznan Street are vague. This picture had a maximum disparity between neighbouring views of 22 pixels close to the Love Bird 18 pixels, so it was expected that it would get a pattern similar to it as well. The reason for this could be content related. Generally are the Poznan Street good compared to the rest of the pictures, but some objects like the traffic sign at the right side of the picture had a lot of crosstalk to it. Some of the subjects may have had focused only on this object and some subjects tried to ignore it. That may have resulted in the vague plots.

When the Z distance is large, the periodic pattern along the X axis exhibits. The reason for this is that the interval between neighbour views becomes larger as measured in section 3.1. The crosstalk are obvious more annoying (blue colour in the plots) at close Z distance. This can be observed form scene content Balloons, Poznan Hall and Book Arrival. The closest 0.5m in the Z direction are the most annoying. This is because the view intervals are so close to each other(less than 4.5cm) that the observed views fuse together and are perceived unpleasant and the amount of crosstalk are also perceived as increasing.

The study of the basic statistics for the scores of the subjective results in table 6.2 and in appendix 3 says something about the crosstalk in general. In this case it easy to see that the Dog picture have less crosstalk then the rest of them based on the score mean and the score distribution plot. Lovebird and Poznan Street are significant better than picture Balloons, Poznan Hall and Book Arrival. The positive aspect of this calculation is very easy to compute and it gives a superficial idea of the amount of crosstalk for each picture. The negative aspects are that it says nothing about the crosstalk related to the position. It just measures the overall perceived crosstalk regardless of position based on subjective scores.

The results and earlier discussion can argue that the subjective crosstalk assessment is consistent with the objective crosstalk measurements that are done before [1, 3]. It is also consistent to the characteristics of an auto-stereoscopic display. The interpolated data plots shows that there is obvious differences among the subjective scores in various scene contents, but the subjective crosstalk assessment is more comprehensive than the objective crosstalk measurement, since the subjective method is also scene content related.

The slice plots in figure 25 and 26 give more details regarding the horizontal distribution. The green curves are not totally smooth. According to theory and objective crosstalk measurements, these curves should be a perfect periodical curve having 6.45 cycles at 2 meters (6.45≈2m/62mm/5). The sine regression used on the raw and interpolated data will approach the perfect periodical curve. It can also be observed that it has 6 or 7 cycles, which is also consistent with the objective measurements.

7.3 ANALYSIS OF MATERIALS AND METHOD USED FOR SUBJECTIVE TESTING

To create a good as possible lab environment, training and test setup was one of the main goals of this assignment. An important criterion was to see if it was possible to derive an inexpensive method to measure crosstalk with subjective opinions. As far as we know, a subjective test that allowed the user to move around in the dedicated area and evaluate the amount of crosstalk from the auto-stereoscopic display has not been done before with success. In addition, the test pictures had different quantity of maximum disparity between neighbouring views, amount and type of content, which affected the result as well. The amount of content relates to how much objects and differences there are in total. The type of content refers to the characteristics, for example the amount of sharp edges, texture and colour composition, depth, contrast etc. In terms of that people perceive the same image content differently; it is very hard to get a perfect result compared to an objective method.

7.3.1 THE DISPLAY

The Tridelity MV4200 was selected since it was new, the largest and probably the best screen available at the department of Electronics and Telecommunications. It is obvious better for the subject to look at a 42" screen at a distance up to 4.5m compared to a smaller one. The 5 view projection was used since the result may be easier to interpret and the creation of a input picture required less work. The multiview screen needs the displaying picture to be in the format showed in figure 9. The training and test pictures were processed using Adobe Photoshop CS4. The result of a 9 view test would be a periodical curve with a lesser frequency compared to the 5 views, granted that the view interval was the same 62mm.

7.3.2 THE HEAD TRACKING SYSTEM.

The Kinect head tracking system worked pretty well. It is an inexpensive method for head tracking. The subject did not need to wear any extra equipment in order to be detected. This was a great advantage compared to the Wiimote system, where the subject needed the helmet to be detected. This was the main reason to go for the Kinect instead of the Wiimote system. The pilot test of the Wiimote system did result in problems with the helmet. The helm position varies for each subject since the head shapes are different. The connection between the helmet and the Wiimote controller was much more sensitive to large movements than the Kinect system. The defined score area could not be larger due to angle restriction of the Kinect.

Other head tracking systems like TrackIR and webcam systems was discarded since these systems have a smaller detection range. The maximum range of these is about 2m. It would be interesting to see an improved Kinect system from Microsoft in action. It is supposed to be way better than Kinect prime sensor. There are rumours that this system can detect face expressions and even lip movements [26]. If this is true, it would be interesting to compare scores to for example face expressions to see if the subjects strive at the position.

7.3.3 SCORE TRACKING

The subjects used a Logitech Bluetooth numpad to type the scores during the test. This worked very well. It was easy for the subject to carry the numpad along since it is wireless and it is intuitive to use. The numpad has buttons from 0 to 9. The 0 button was used under calibration and buttons 1-5 was used during the test. The buttons from 6-9 were spare button, which were not used. If a subject pressed one of the buttons from 6-9 or the 0 button during the test, it got misprinted with a score tag of -1 in the data file. All the -1 scores in the file had to be removed or ignored while processing the data in mat.lab.

7.3.4 SCENE CONTENT

The scene content had definitely impact on the result. If we compare the data plots above in figure 25 and 26, there are obvious differences. Regardless of the content, the pictures' maximum disparity from neighbouring views had the most impact on the result. A high disparity resulted in low scores, but still provoked differences in the picture.

It would be interesting to have one picture with one view displaying a totally white colour and the rest of the views turned black to see if the same patters occurred regardless of the content.

7.3.5 LAB ENVIRONMENT

The restricted moving area worked well. The resulting surface made from the data gave a pretty good basis for comparison with an objective expectation. The angle limitations of the Kinect system became also the limitation of the moving area. This area could have been smaller to get more data points per square meter, a more focused area. Less movement would also be an effect of this. That could lead to a better result since the outliers would have had less impact on the result. These assumptions given that we still had 25 test subjects.

The movement while sitting in the chair resulted in that the Y position altered, with 33.6mm in average. Since the amount of crosstalk perceived changes with the height it affected the result more or less. It is hard to estimate or correct for this alternation in Y position. The Kinect system receives the Y position in a non linear way depending on the distance from the sensor, the Z position.

It also occurred that the lowest subjects had some problems to move while sitting still as possible since their feet barely touched the floor. It resulted in bigger physical movements to change their position which way have had bad influence on the result as well. Ideally, the subjects should have had the possibility to automatically move the chair around in any direction always facing the screen. This would require a special made chair which would violate the purpose of an inexpensive methodology.

The light in the room was good. Just a little unintentional light came from the windows, but I think this was so little that it did not affect the result.

7.3.6 THE SUBJECTS AND VISUAL SCREENING

The subjects reacted quite similar to the crosstalk. As expected was the Balloons, Poznan Hall and Book Arrival pictures a little strenuous. These three pictures had the worst disparity, respectively 30, 38 and 40 pixels. Only one subject had to abort the test session after a while due to a little nausea. This subject was also pregnant which could be an additional factor to this nausea. Two of the subjects were a little colour blind. Both failed on one of five colour trial before the test. As this occurred they were also asked to point out specific crosstalk on the training pictures in order to proceed to the test.

Every subject that participated had a stereo vision smaller than 30 arc second, but still there were some small differences. Every subject said they could perceive the depth in the pictures. The level of perceived depth and how the crosstalk interference affects the subject will always differ. Some subjects are more sceptical in their judgment when scoring the pictures despite the stimuli that was done before the test. These variations will affect the result.

7.3.7 THE TRAINING SESSION

The training session worked pretty well. All the subjects got an idea of the scores, but still it seems that the subjects found it hard to distinguish between score 3 and 4. The result could improve if the training session had an additional training picture with scores between 3 and 4. The session lasted for about 5-6 minutes for each subject. The result could have been better if the subjects got a longer practise time as well. In theory could the subject use as long time as they desired, but they seemed eagerly to start the test.

Another improvement could have been a test system based on objective measures captured in same conditions as the test. The test subject could have rated a picture in some positions and the system could validate the scores based on the objective measurements. So if the typed score in a certain position was consistent with the objective prediction, the subject had proven that they had the same understanding as the system. A system like this would have required a lot of work and would be unrealistic before this test, but it could work to improve the training.

7.3.7 THE TEST SESSION

Since this test utilizes single stimuli the subjects had to memorise the reference pictures and compare five different score to each other. Some subjects uttered that they rated the first and maybe the second test picture too strictly or too gentle compared to the rest of the pictures. This imbalance may be prevented since each subject evaluated the pictures in a random order. So statistically should this imbalance be uniformly distributed, but since it was only 25 test subjects could this distribution be some skewed in either direction.

The test took about 45 minutes to complete. This may have been too long. Some subjects said they got a little tired at the end of the session. It would be interesting to see if the result got better if we only used 5 pictures instead of 6, but this is just speculations.

Objective part

7.4 GOAL AND EXPECTATIONS OF OBJECTIVE PART

The main objective of this part was to create an objective representation of the crosstalk influenced pictures that was observed during the subjective crosstalk experiment. A Pearson correlation function was applied to find the best as possible correlation between the subjective results and objective results.

It is expected that it is complex to create a fully credible representation of a subjective perceived crosstalk influenced picture. Crosstalk is complex and is often perceived different from a subject observation compared to an objective representation. It is expected that the methods created and used in this experiment can be improved a lot. We will discuss more around thoughts and suggestions for improvement in the coming sections.

7.5 CREATING CROSSTALK-DISTORTED PICTURES

There are done several simplifications in order to create an objective representation of the pictures that was utilized during the subjective experiment. The influence of crosstalk is based on the neighbour that influences the current view the most. The result may have been better if the contribution from every view was taken into account, but this makes it much more complex. It further work, it would be interesting to see the difference in correlation between the simplification and the more complex approach. This experiment focused more on if the subjective results could correlate to an objective measure based on fixed positions at all.

The creation of objective crosstalk images are based on the transparency of the captured black and white pictures. Less transparent gives more added crosstalk from another view. It is also possible to determent the intensity of crosstalk based on a grayscale threshold value, which could also provide a interesting result.

An additional approach would be to render the influenced pixels with a percentage that was depended of the difference of the reference colour and the captured black and white image, this method are derived in [4]. This method can be summarized with function 5 below.

$$\begin{cases} R_c^p = R_o + p \times L_o \\ L_c^p = L_o + p \times R_o \end{cases} \quad (5)$$

The new right and left side crosstalk influenced pictures equals the right and left original images, Ro and Lo, summarized with a p value that indicates the percentage of added crosstalk. It would result in a normalized value between 0 and 1, which tells how much a pixel was disturbed by the crosstalk [18]. This would result in a more noisy picture that could have given a better correlation with the subjective results.

7.6 EXTRACTION OF SUBJECTIVE VALUES AND INTERPOLATION

The extraction of subjective values was one of the main challenges, especially since this experiment is quite unique. The extraction that was done was based on the model that was derived in section 5.3. The accuracy of the extraction could be improved by filter the sine period to a more rectangular form. One way that may work is to adapt the sine wave to fit Gibbs phenomenon for a sine wave [7]. This new wave would fit the model better and may generate a better correlation with the objective values. Using a support vector machine (SVM) could help to determine the decision boundaries between the 4L and L areas in model derived in section 5.3 better.

It could be debatable if it would be an interesting direction to study more than one period of the view pattern, like three or four of the periods. I think the result would be almost the same and it would require a lot of camera captures of the screen. The image pre-processing would also take a lot of time. The 4L – L ratio derived figure 24 may change a little with increasing viewing angle. This will affect the accuracy of the sampling and must be taken into account.

The tile plots in figure 29 did not work, since the density of the user scores was too low and resulted in a bad correlation. The only information these plots give that the first impression says that crosstalk is non-uniform. The amount of crosstalk relates to the position and it has a periodically pattern. The probability of receiving a score of 4 or 5 in the yellow and red areas is above 80%. It states only that you with a high probability perceive less crosstalk in these areas than in “colder” positions according to the heat diagram. This method could also be adapted to find the 10 closes neighbours instead of tiles, which are limited by borders.

7.7 ANALYSIS OF CORRELATION RESULTS

The SSIM correlation of the pure sine wave varies from best 0.79 for picture Lovebird to 0.25 for Poznan Hall. The reasons for this variation is related to the content in the picture were the crosstalk applies, besides from how the objective picture was created and the method for extraction of subjective values. Using only the sine wave to sample subjective values gives a poor correlation for most of the pictures, since the sine curve have a uniform frequency that do not correspond to the model in section 5.3.

The PSNR reflects the SSIM pretty well except for when all pictures are taken into account. The overall result for all the pictures taken into account is not a realistic measure. If you observe the average scores of the Dog and Book Arrival pictures , which is 3.70 and 1.57, and observe the min and max score of the SSIM and PSNR values respectively 0,994519 – 0,999198 and 40,9217- 50,1145 for the Dog picture and 0,978815 – 0,992903 and 29,4941 – 36,3201 for the Book Arrival picture. The important clue is that the best SSIM values of Dog and Book Arrival are too close. In reality is the quality of these pictures far away from each other. This argument sustains the fact that the objective method for creating crosstalk influenced images is not complex enough to compare all the images against the same range of SSIM values. The method seems good for rating the quality of pictures disparity values below 38 pixels.

The result of using both the sine wave and the linear interpolation gives a much better correlation for single pictures, except for Poznan Street. The Poznan Street is a special case picture that will be derived in section 7.7.1 below. The SSIM correlation increases with values between 0,065-0,3845 for rest of the pictures. This can be viewed in table 5. It indicates that the model derived in section 5.3 is a step in the right direction. It is plausible that this correlation will even improve with a better extraction that mimics the model in section 5.3.

Another way of improving the results is to parameterize the SSIM values with function 4. The resulting correlation increases with values between 0,016 – 0,1601. Both the pictures Dog and Lovebird have a high correlation (> 0,8981). These pictures have a low base disparity in common, respectively 8 and 18 pixels. The method that was used for creating objective crosstalk pictures combined with the extraction model seems to favouring pictures with a low disparity. This appears reasonable since the SSIM values for expected good quality pictures like Dog and Lovebird and expected poor pictures seems to not differ enough. The pictures of the worst case scenario in appendix 4 underpin the statement above.

The correlation result is good if Poznan Hall and Book Arrival are excluded when the common correlation value is calculated for all the pictures together. ***The correlation is 0,8249 when we use the values extracted with the combined sine and linear interpolation method together with parameterized SSIM values. This result is the most important for the whole experiment and it proves that the way subjective values are extracted and both the creation and calculation of objective material is quite good***. The methods can be improved a lot in order to get an even better correlation, especially when pictures with high disparity are included.

7.7.1 SPECIAL CASE: POZNAN STREET

The start value during extraction of subjective values was fixed to the lowest value of the sine wave period. This extracted value was mapped to the assumed worst SSIM value, which was the transition between view 1 and view 5. If you note the value of the Poznan Street in table 3, it is not the worst value. The worst SSIM value appears in the transition between V1 and V2. The content placement in this picture influences the SSIM value more between V1 and V2 than the expected influence between V1 and V5, which applies for the rest of the test images. Content that are closer to the camera will be more influenced by crosstalk for the observer[1]. So if a object in the picture appears in the foreground in only a fraction of the views, it may cause a greater influence than the expected disparity that occur in the transition area between V1 and V5. In the Poznan Street picture, the red stop sign to the right side do not appear in view 5 and partly view 4. This sign affect the first views more than the transition area betweenV1 and V5.

If the transition area between view 1 and 2 is mapped to the first value of the subjective extraction for the Poznan Street picture, the SSIM correlation improves from 0.6558 to 0.8293 using the sine wave only to extract subjective values. By using the linear interpolation as well the SSIM correlation result improve from 0.8293 to 0.9499.The correlation ends on 0,9762 by parameterize the SSIM values.

The conclusion of this special case is that the worst value of the extraction should always be mapped to the worst value of the SSIM value regardless of the view pair. The rest of the sampling should continue with the same samplings frequency.

7: CONCLUSION

This experiment has proposed a new and realistic subjective crosstalk assessment methodology on auto-stereoscopic display with consideration of different scene contents and viewing positions. The subjective results show their consistence to the characteristics of auto-stereoscopic display, while provide more detailed crosstalk perception information in terms of viewing position and scene content when compared to the objective crosstalk measurement. The resulting surfaces can be used to design a crosstalk perception metrics for auto-stereoscopic displays. The methodology can be improved in many ways. Some improvements can be; better head tracking system, improved movement system, decreasing the inconsistent between subjects by improving the training, increasing the amount of subjects, post processing of data and find a good method for removing outliers. The method for objective creation of crosstalk influenced pictures works good for low and medium disparity pictures when the minimum subjective value are mapped to the minimum SSIM value. The correlation are at the highest 0,97. When all the six test pictures are taken into account deteriorates the correlation a lot. The correlation ends at 0,5139. The main reason for this is that the mapping of SSIM values to subjective values is not linear when all the pictures are taken into account. One way to solve this is to create a more complex objective crosstalk influenced picture which represent the actual quality better. When the two pictures with the highest disparity are removed from the correlation it improves a lot. The final correlation is 0,8249 which gives a good indication that the utilized methods are quite good for pictures that have a maximum disparity below 38 pixels. The correlation provide also an indication of the quality related to the subjective test methodology.

7: REFERENCES

[1] Barkowsky, Marcus., Campisi, Patrizio., Le Callet, Patrick,. & Rizzo, Vito.(2010) *"Crosstalk measurement and mitigation for autostereoscopic displays"*. Department of Applied Electronics, University "Roma Tre". Italy.

[2] van Berkel, C. (1999) *"Image preparation for 3D-LCD,"* in Proceedings of SPIE, 1999, vol. 3639, pp.84-91.

[3] Boev, Atanas., Gotchev, Atanas., & Egiazarian, Karen. (2007). *"Crosstalk measurement methodology for auto-stereoscipic screens"*.3DTV-CON 2007. Institute of signal Processing, Tampere University of Technology, Tampere, Finland.

[4] Boev, Atanas., Gotchev, Atanas & Hollosi D. (2009) *" Mobile3DTV Project report, no 216503".* Institute of signal Processing, Tampere University of Technology, Tampere, Finland.

[5] Dodgson, Neil A. (2011). Professor of Graphics & Imaging.*"Multi-view autostereoscopic 3D display"*. University of Cambridge.

[6] Dodgson, Neil A. (2002). *"Analysis of the viewing zone of multi-view autostereoscopic Displays"*. University of Cambridge Computer Laboratory, Cambridge, UK.

[7] Hewitt, Edwin; Hewitt, Robert E. (1979). "The Gibbs-Wilbraham phenomenon: An episode in fourier analysis". *Archive for History of Exact Sciences* **21** (2): 129–160.

[8] Jain, A., & Konrad, J. (2007) *"Crosstalk in automultiscopic 3-D displays: Blessing in disguise?"* SPIE Symposium on electronic Imaging, Stereoscopic Displays and Applications, San Jose, CA, USA.

[9] Konrad, J., & Agniel, P.(2004). *"Non-orthogonal sub-sampling and anti-alias filtering for multiscopic 3-D displays"* in Proc. SPIE Stereoscopic Displays and Virtual Reality Systems, vol. 5291, pp. 105-116.

[10] Konrad, J., & Agniel, P. (2003). *"Artifact reduction in lenticular multiscopic 3-D displays by means of anti-alias filtering"* in Proc.SPIE Stereoscopic Displays and Virtual Reality Systems, vol. 5006A, pp. 336-347.

[11] Lipton, L. (1987) *"Factors affecting 'ghosting' in time-multiplexed plano-stereoscopic CRT display systems".* True 3D Imaging Techniques and Display Technologies, vol. 761, pp. 75-78.

[12] Loza et al., "Structural Similarity-Based Object Tracking in Video Sequences", Proc. of the 9th International Conf. on Information Fusion.

[13] Onural, L., Sikora, T., Ostermann, J., Smolic, A., Civanlarand, M. R., & Watson, J. (2006) *"An Assessment of 3DTV Technologies,"* NAB Broadcast Engineering Conference Proceedings 2006, pp. 456-467, Las Vegas, USA.

[14] Pastoor, S. (1995) *"Human factors of 3D images: Results of recent research at Heinrich-Hertz-Institut Berlin".* International Display Workshop, vol. 3, pp. 69-72.

[15] Rahman, N.A. (1968) *A Course in Theoretical Statistics*, Charles Griffin and Company

[16] Seuntiens, P. J. H., Meesters, L. M. J., & IJsselsteijn, W. A. (2005*)"Perceptual attributes of crosstalk in 3D images"*. Displays, vol. 26, no. 4-5, pp. 177-183.

[17] Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004) "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612.

[18] Xing, L., Ebrahimi T., & Perkis, A. (2010) *"Subjective evaluation of stereoscopic crosstalk perception". Visual Communications and Image Processing, vol. 7744, Huang Shan, China.*

[19] ISO/IEC JTC1/SC29/WG11, M15377, M15378, M15413, M15419, Archamps, France, 2008.

[20] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation, BT.500-11*, 2002

[21] Coding4fun. (2009). Obtained 01.12.2011 at 21.10 from http://channel9.msdn.com/coding4fun/articles/Wiimote-Virtual-Reality-Desktop

[22] Drakos, N. (2002). "*Cubic spine Interpolation*". Obtained 02.12.2011 at 13.32, from http://www.physics.utah.edu/~detar/phys6720/handouts/cubic_spline/cubic_spline/node1.html

[23] Gale, T .(2006). "*Fundamental Mathematical Concepts and Terms*". Obtained 15.12.2011 at 03.11, from http://www.novelguide.com/a/discover/rlm_0001_0002_0/rlm_0001_0002_0_00075.html

[24] Szasz, T., Katai-Urban, G., Petrovanster, B., & Pereanez, M. (2011). Obtained 08.12.2011 at 19.52, from http://www.inf.u-szeged.hu/projectdirs/ssip2011/teamA/hardware.html

[25] Lee,J.C. (2008). Obtained 16.12.2011 at 09.42, from http://www.cs.cmu.edu/~15-821/CDROM/PAPERS/lee2008.pdf

[26] Lorentzen, J. (2011). Obtained 17.12.2011 at 16.47 from http://www.spill.no/default.aspx?section=artikkel&id=2204

[27] Paul. (2011). *"What is crosstalk?*". Obtained 06.12.2011 at 17..22 from http://www.tvinformatie.com/wat-is-crosstalk-ghosting

[28] PrimeSense Inc. (2010). Obtained 22.11.2011 at 12.55 from http://andrebaltazar.files.wordpress.com/2011/02/nite-controls-1-3-programmers-guide.pdf

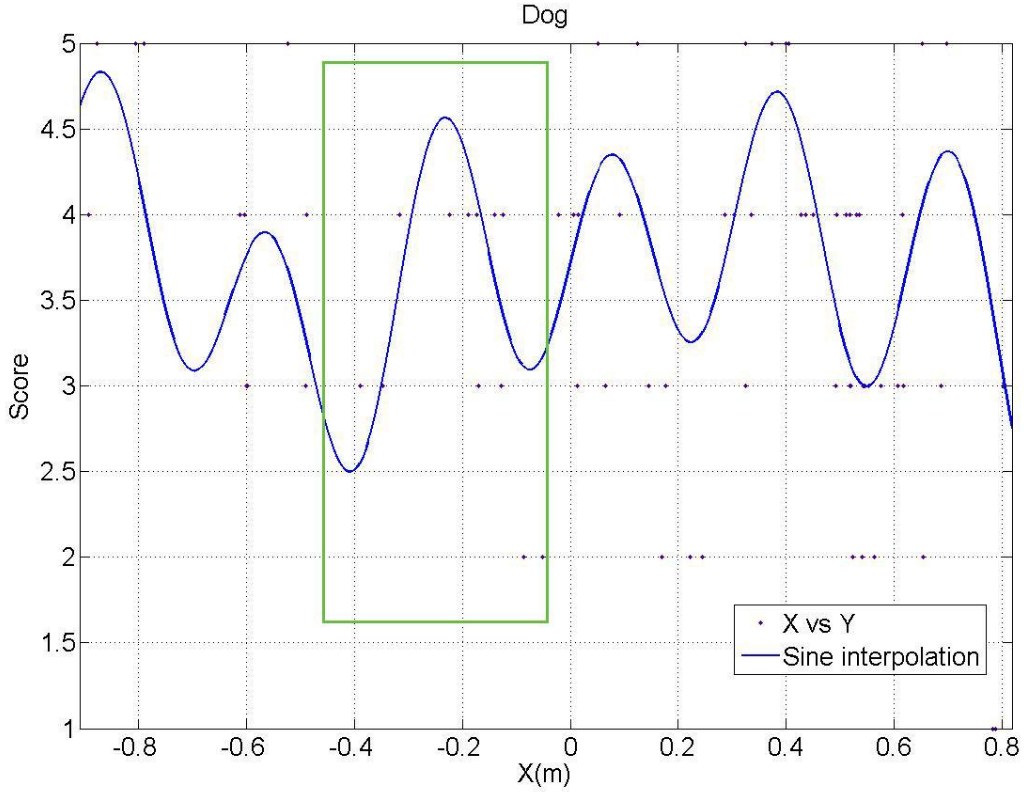
[29] MeTriX MuX Visual Quality Assessment Package. Obtained 11.04.2012 http://foulard.ece.cornell.edu/gaubatz/metrix_mux/

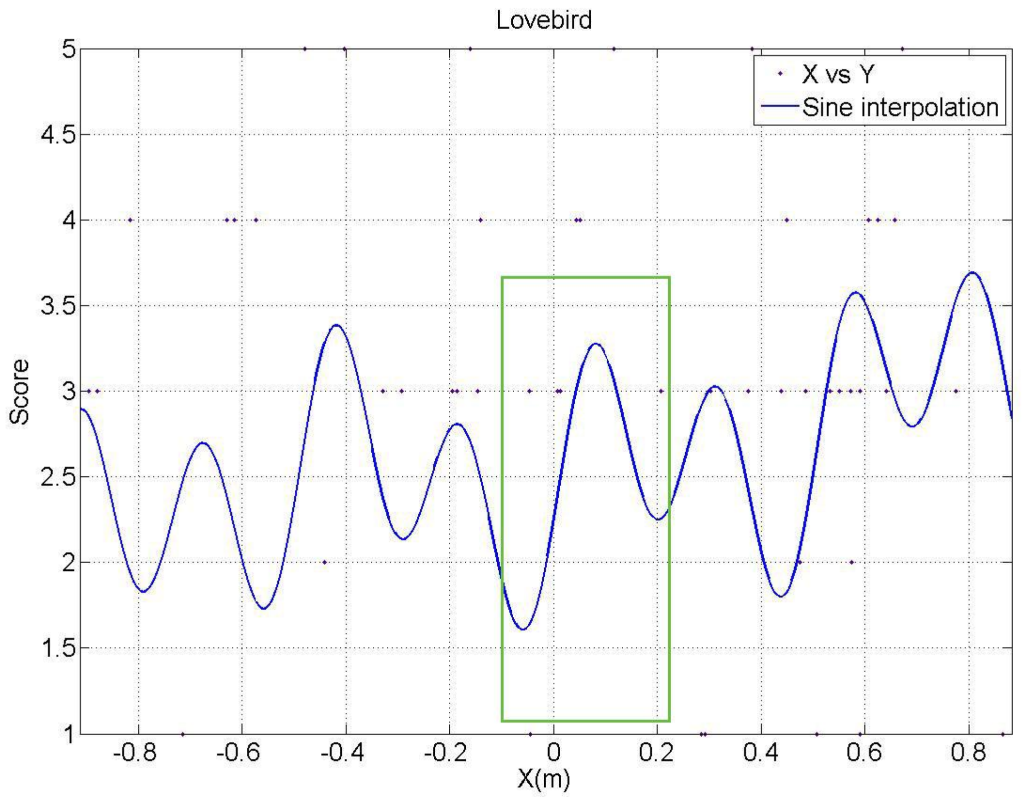
[30] Soft Service company. (2011). Obtained 15.12.2011 at 22.10 http://www.btframework.com/wiimote.htm

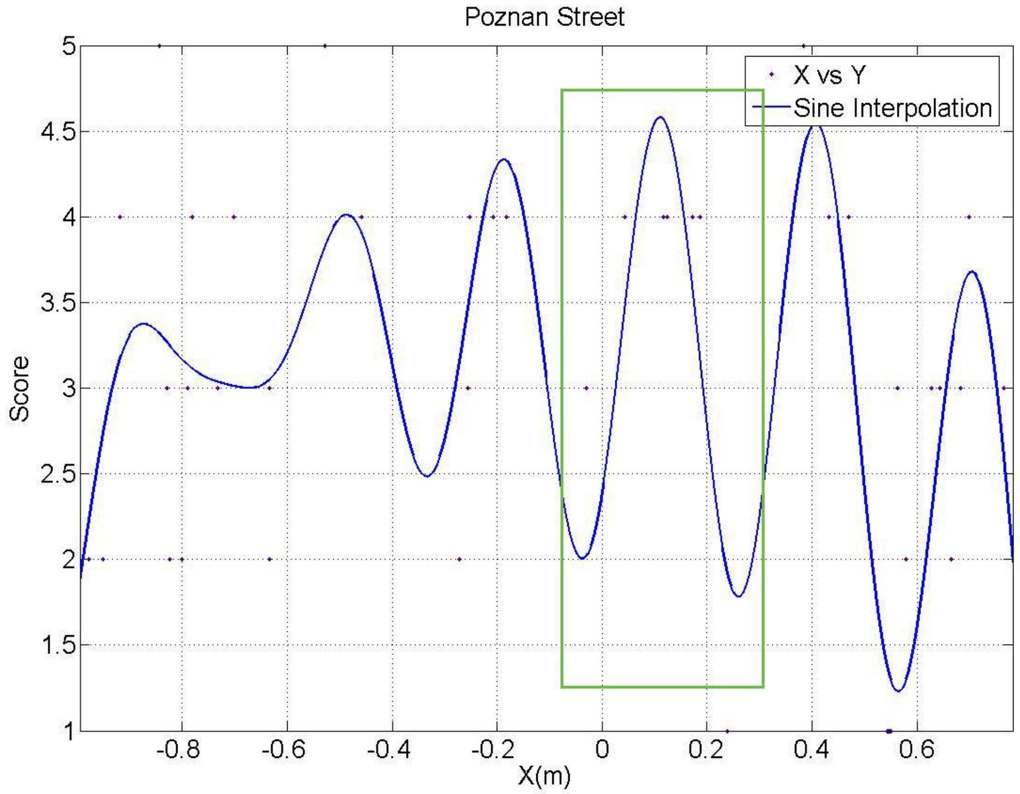
[31] Tridelity. (2011). Obtained 15.12.2011, from http://www.tridelity.com/3D-Display-MV4210va.3d-display-mv4200.0.html

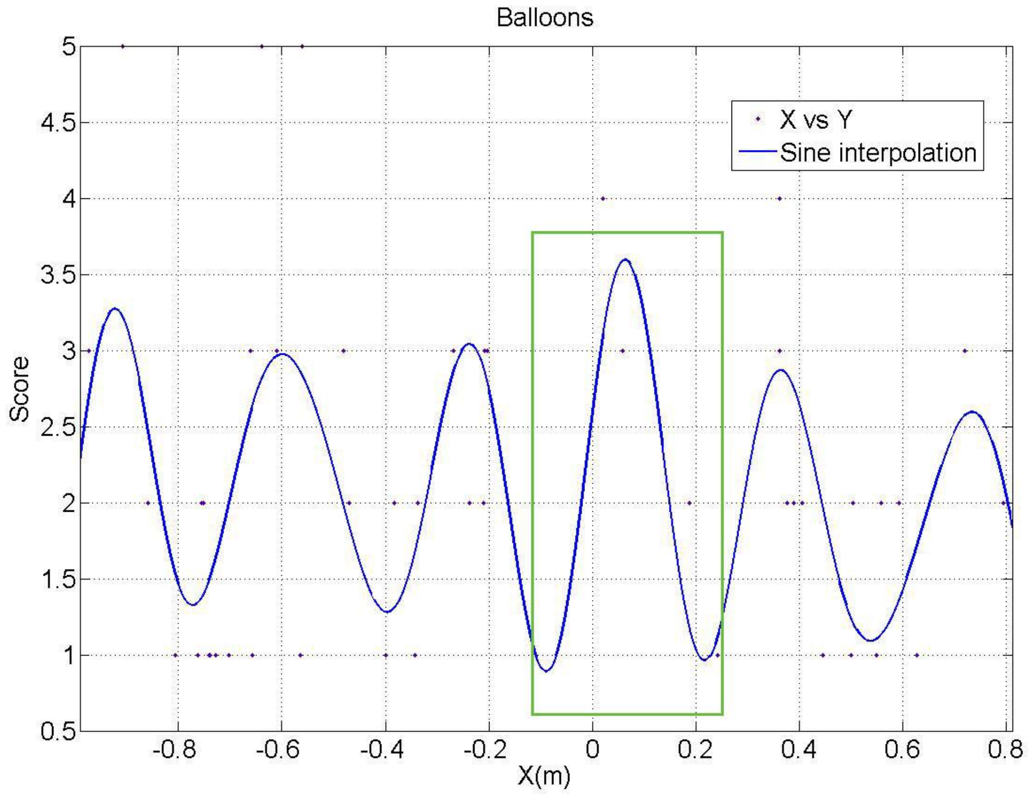
[32] ITU-R BT. 500-11 *"Methodology for the subject assessment of the quality of television pictures" (1974-2002).*

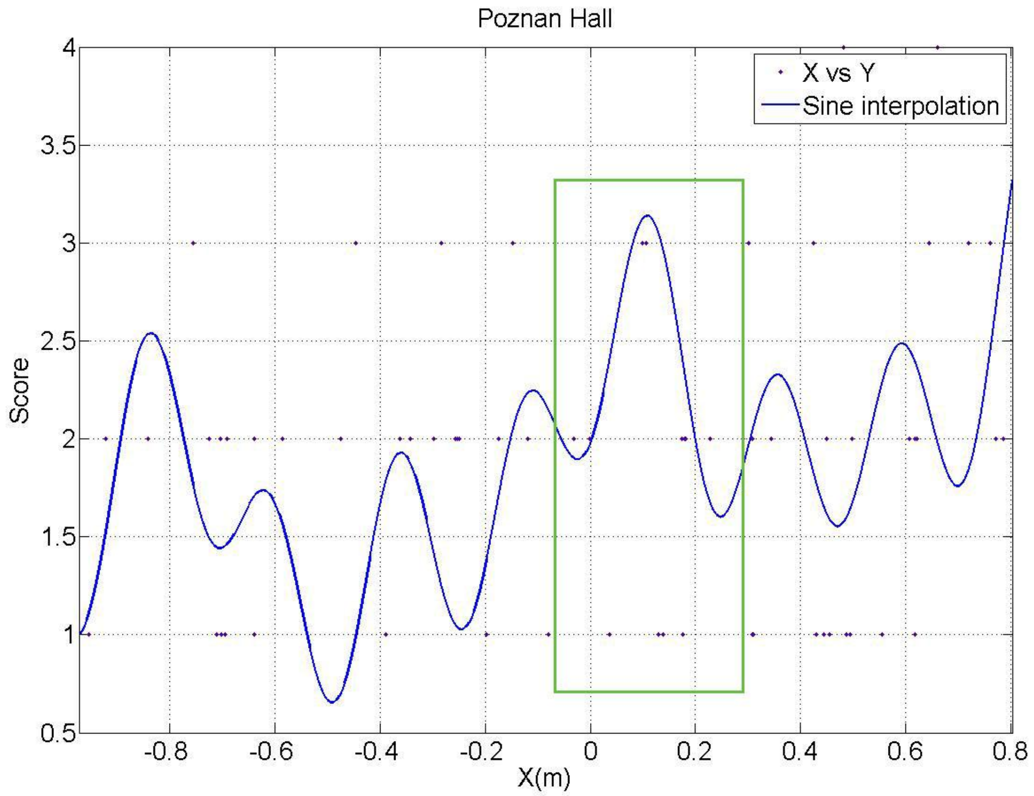
Appendix 1: Sine interpolation plots with marked sampling area.

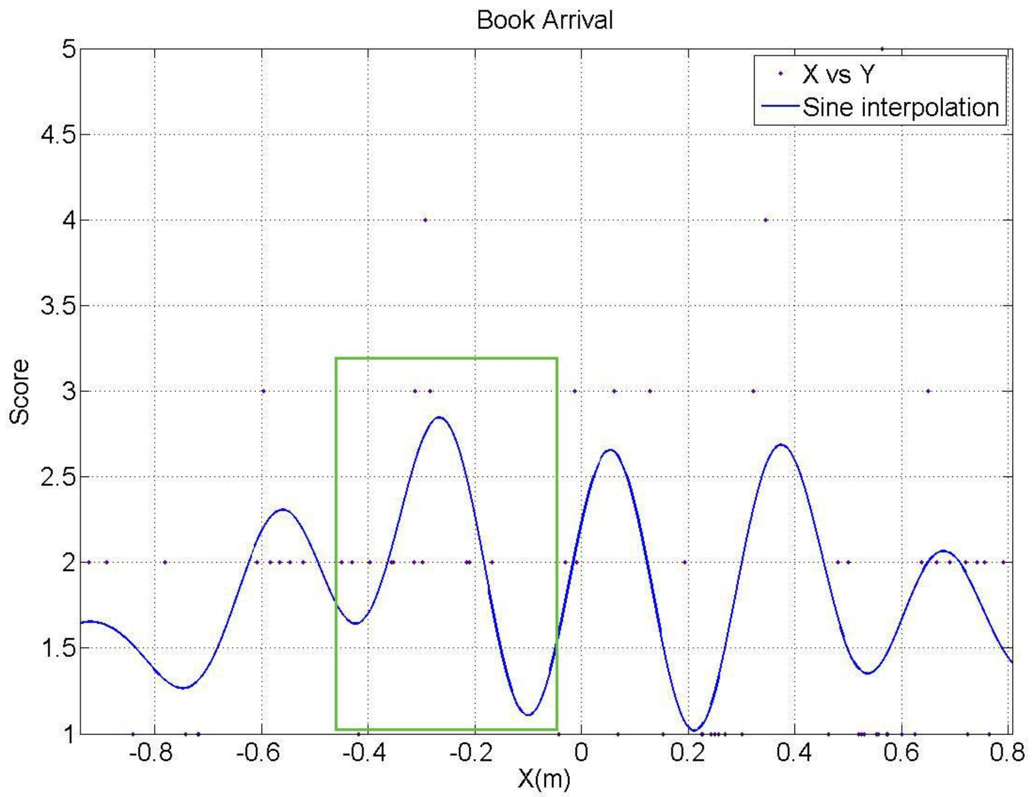




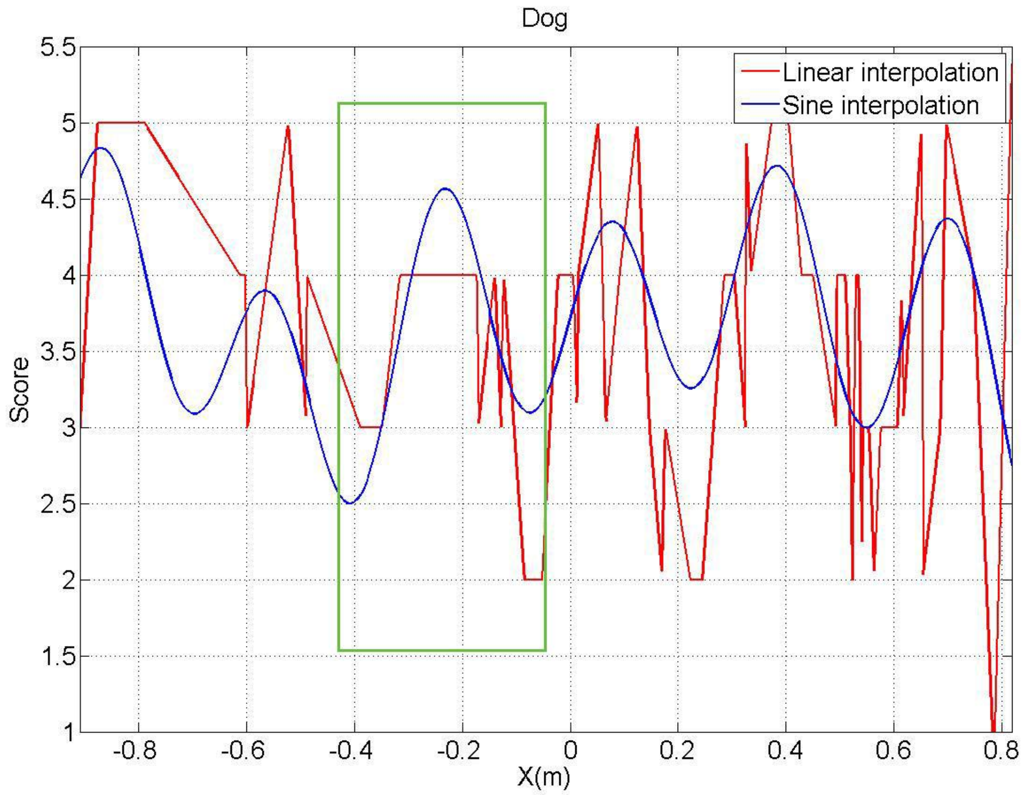

Poznan Street
X vs Y
Sine Interpolation
Score
X(m)

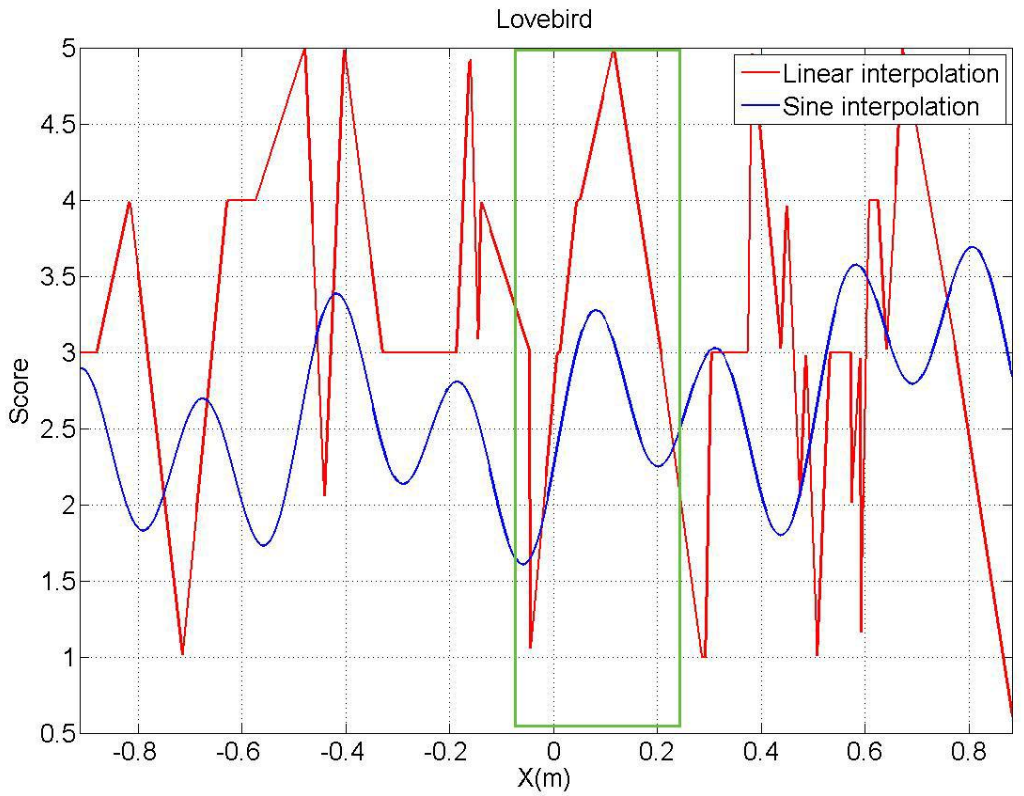

Balloons
X vs Y
Sine interpolation
Score
X(m)

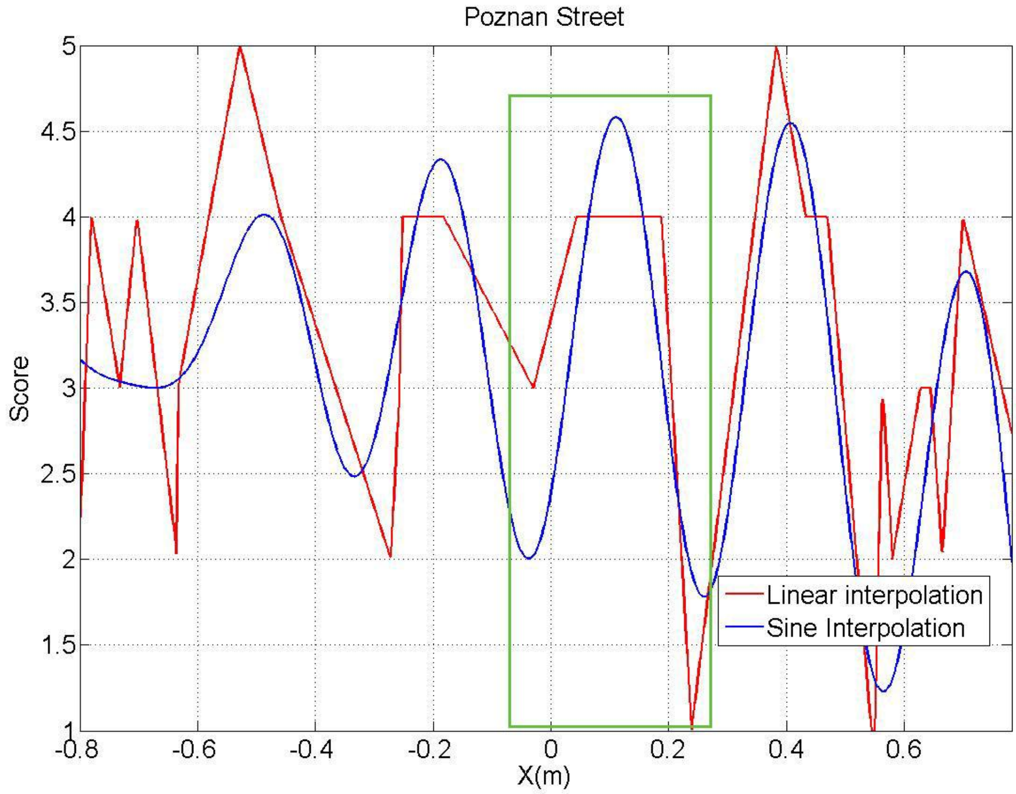

Poznan Hall
4
3.5
3
2.5
2
1.5
1
0.5
Score
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
X(m)
X vs Y
Sine interpolation

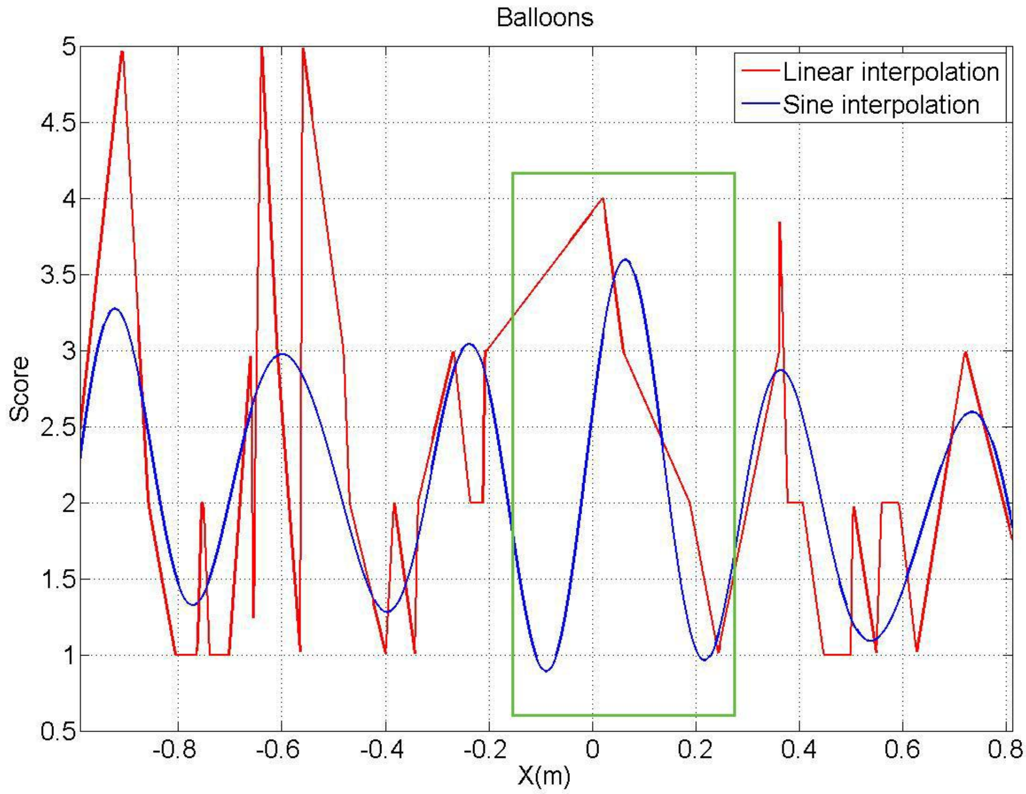

Book Arrival
5
4.5
4
3.5
3
2.5
2
1.5
1
Score
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
X(m)
X vs Y
Sine interpolation

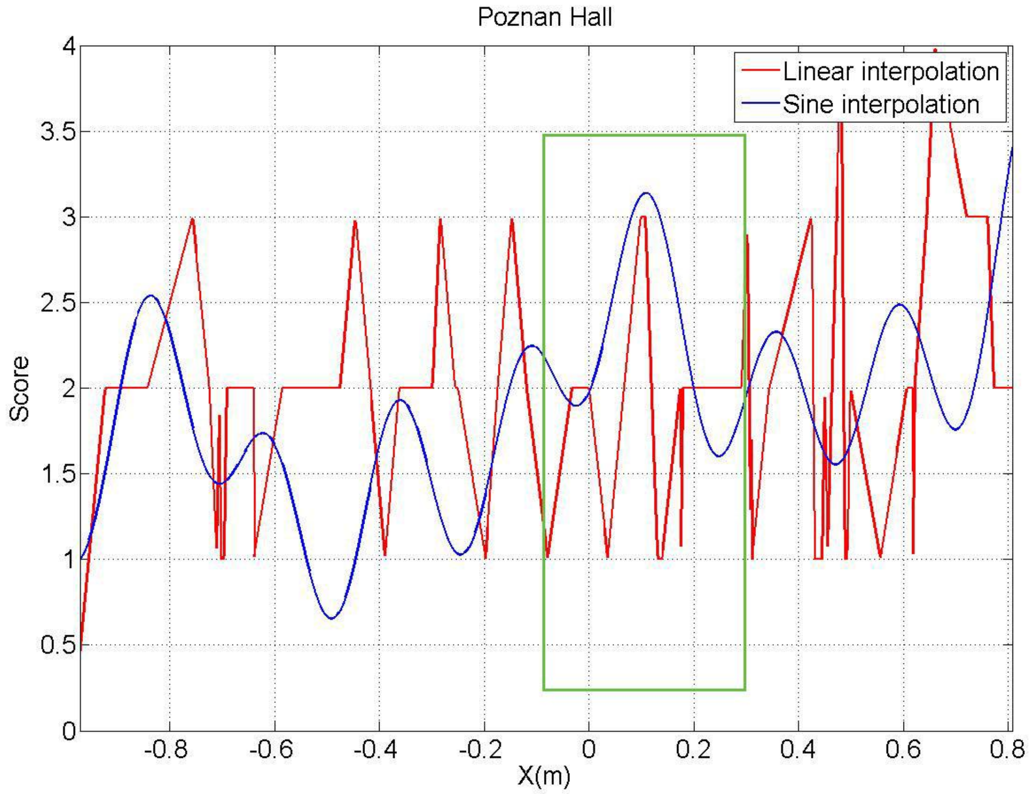
Appendix 2: Combined sine and linear interpolation plots with marked sample area

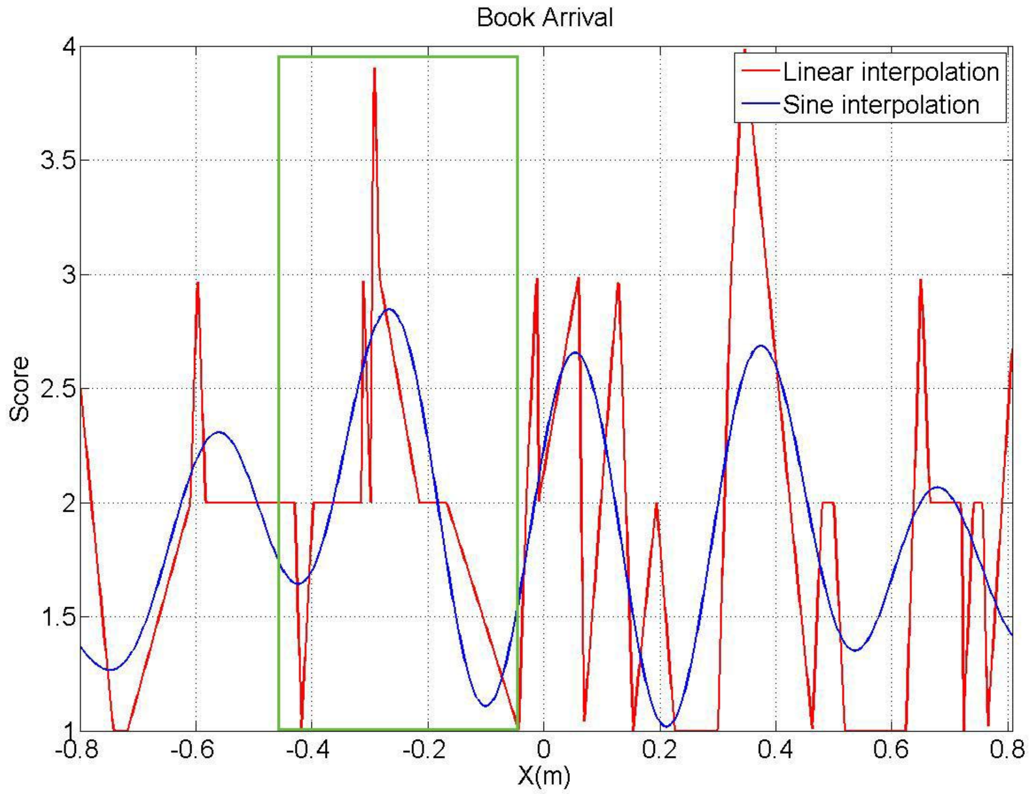





Poznan Street
Score
X(m)
Linear interpolation
Sine Interpolation

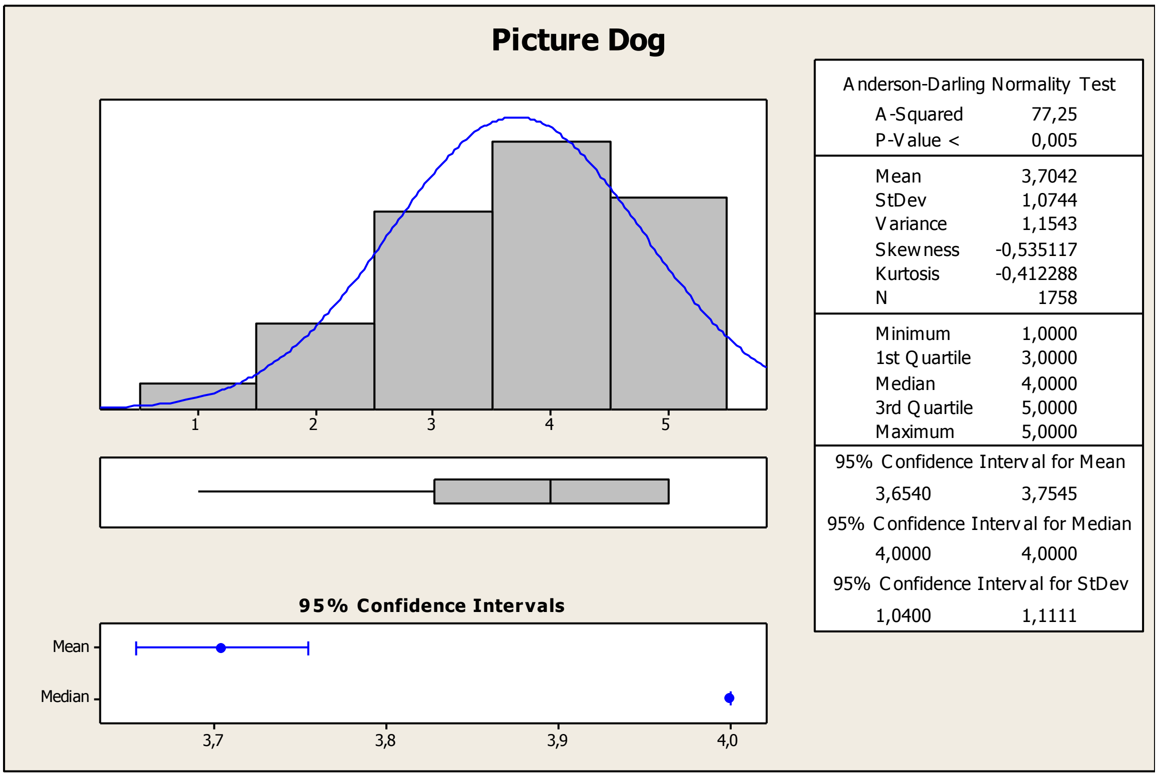

Balloons
Score
X(m)
Linear interpolation
Sine interpolation


Poznan Hall
Linear interpolation
Sine interpolation
Score
X(m)
0
0.5
1
1.5
2
2.5
3
3.5
4
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8

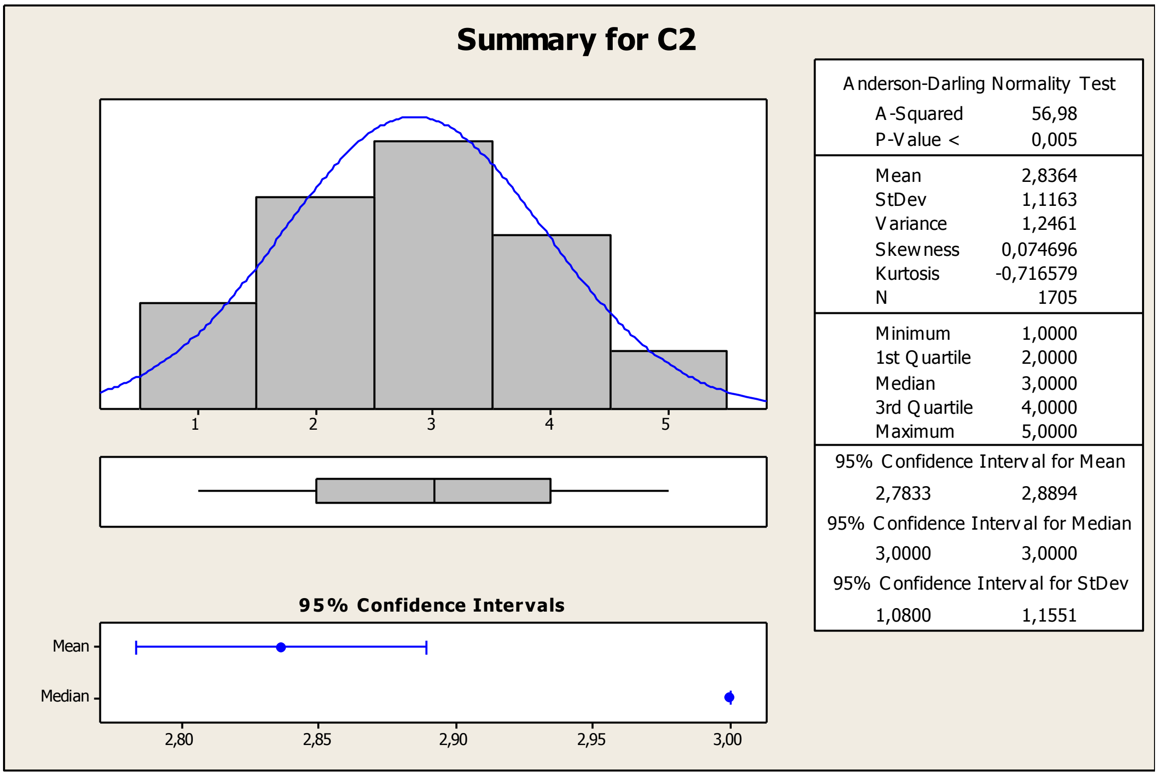

Book Arrival
Linear interpolation
Sine interpolation
Score
X(m)
1
1.5
2
2.5
3
3.5
4
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8

Appendix 3: Basic statistics for the subjective results for each picture

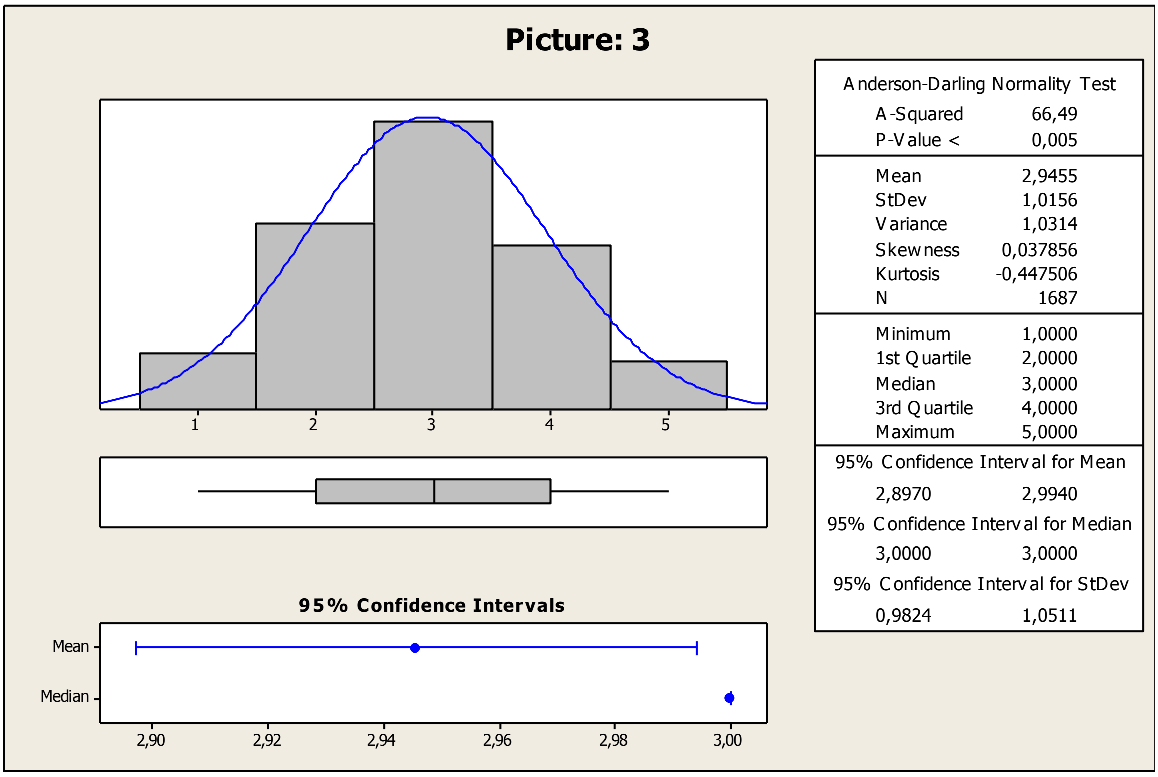
Picture1:



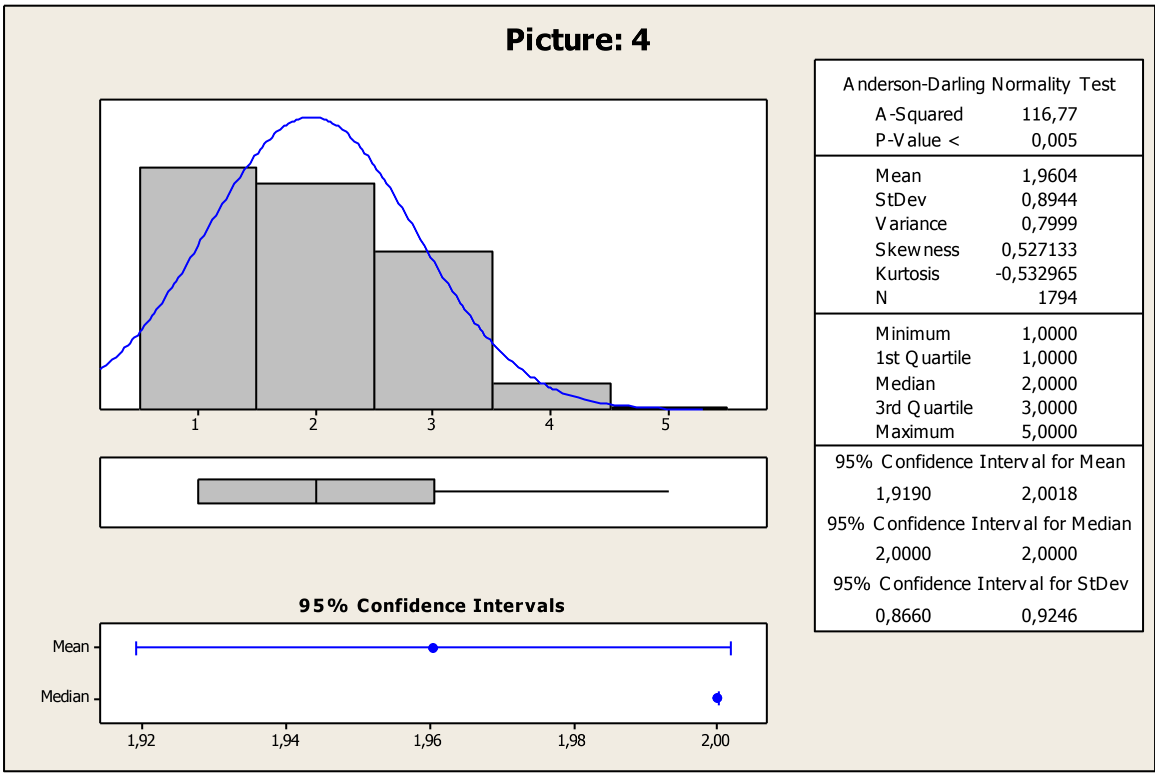
Picture2:



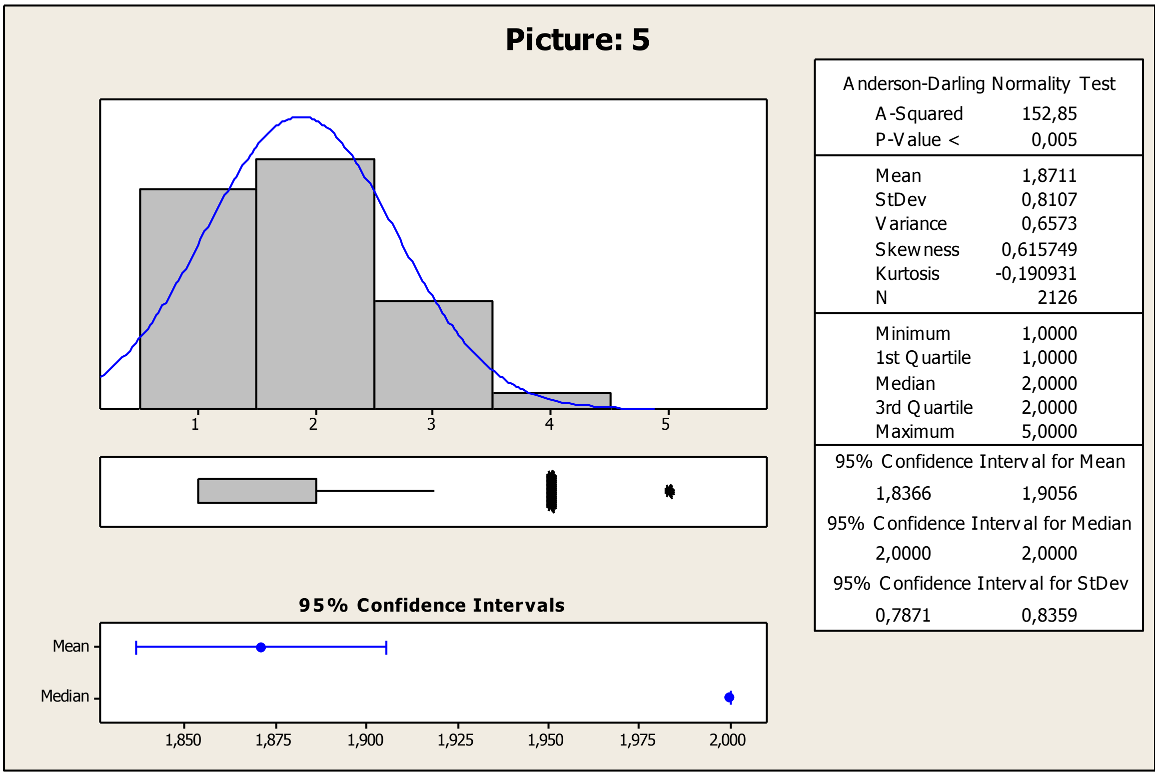
Picture3:



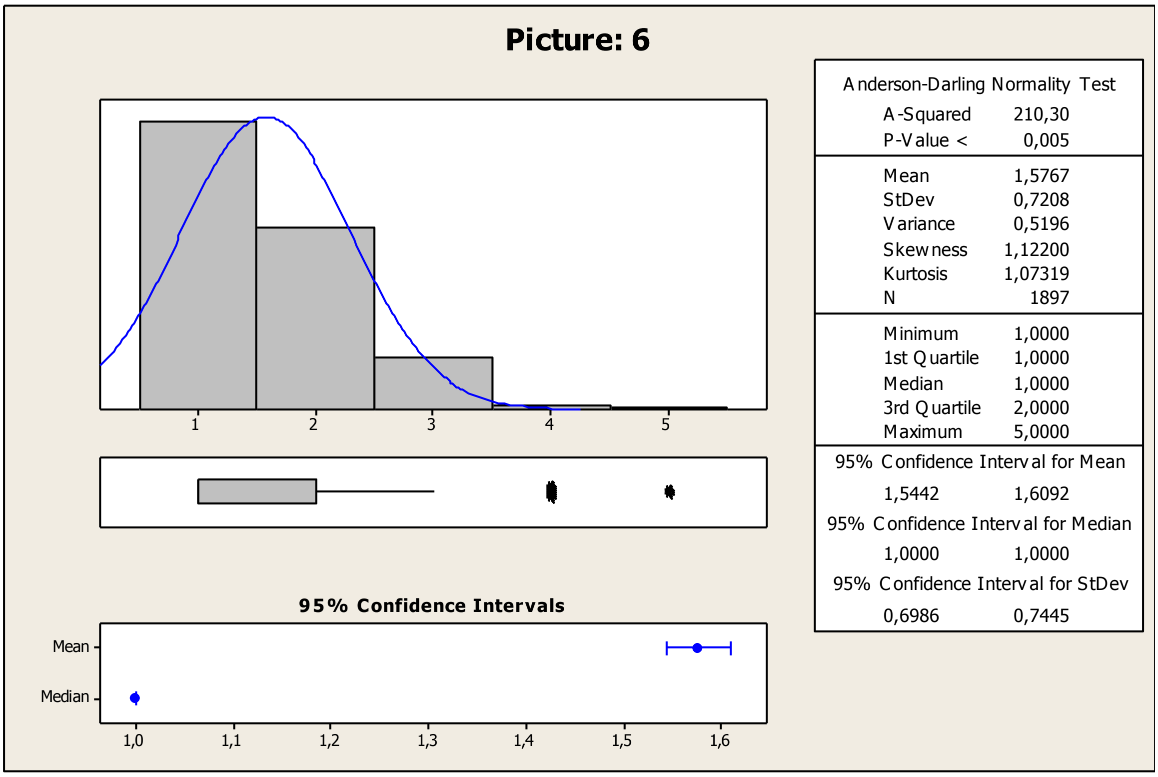
Picture4:



Picture5:



Picture6:



Appendix 4: Most influenced crosstalk pictures.

Dog

Lovebird

Poznan Street

Balloons

Poznan Hall

Book Arrival